

AD-A176 672

THE FEASIBILITY OF USING PYROLYSIS-MASS SPECTROMETRY
AND PYROLYSIS-MS/MS. (U) COLORADO SCHOOL OF MINES
GOLDEN CO DEPT OF CHEMISTRY-GEOCHEMI..

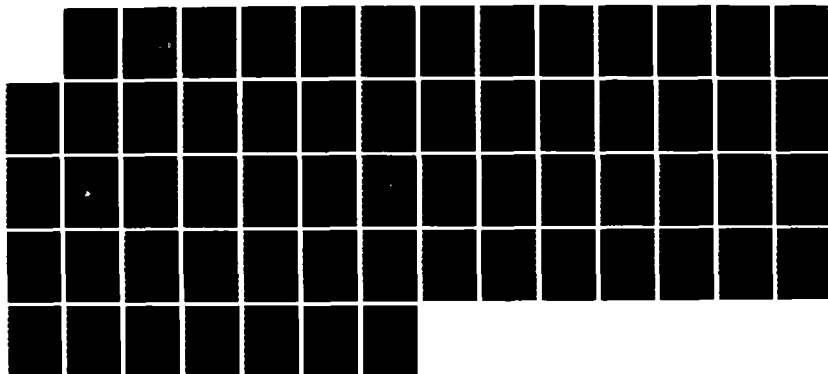
1/1

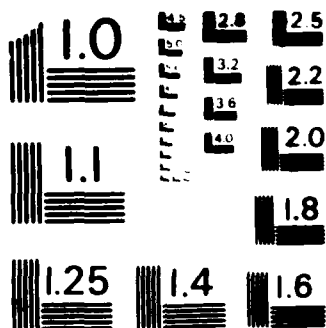
UNCLASSIFIED

K J VOORHEES ET AL. 07 JAN 87

F/G 7/4

ML





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

②

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM										
1. REPORT NUMBER ARO 22961.1-LS	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A										
4. TITLE (and Subtitle) "The Feasibility of Using Py-MS and Py-MS/MS with Pattern Recognition for the Identification of Biological Materials"		5. TYPE OF REPORT & PERIOD COVERED Final 8/1/85 - 9/30/86										
		6. PERFORMING ORG. REPORT NUMBER										
AUTHOR(s) Kent J. Voorhees Steven L. Durfee		8. CONTRACT OR GRANT NUMBER(s) #DAAG29-85-K-0199										
PERFORMING ORGANIZATION NAME AND ADDRESS Department of Chemistry and Geochemistry Colorado School of Mines, Golden, CO 80401		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS										
CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE 1/7/87										
7. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 56										
		15. SECURITY CLASS. (of this report) Unclassified										
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE										
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.												
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) NA												
18. SUPPLEMENTARY NOTES The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.												
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)												
<table border="0"> <tr> <td>Pyrolysis/mass spectrometry</td> <td>Mixture analysis</td> </tr> <tr> <td>Pyrolysis/tandem mass spectrometry</td> <td>Pollen</td> </tr> <tr> <td>Pattern Recognition</td> <td>Atmospheric Particles</td> </tr> <tr> <td>Bacteria</td> <td>Fungi</td> </tr> <tr> <td>Pyrolysis-MS/MS</td> <td></td> </tr> </table>			Pyrolysis/mass spectrometry	Mixture analysis	Pyrolysis/tandem mass spectrometry	Pollen	Pattern Recognition	Atmospheric Particles	Bacteria	Fungi	Pyrolysis-MS/MS	
Pyrolysis/mass spectrometry	Mixture analysis											
Pyrolysis/tandem mass spectrometry	Pollen											
Pattern Recognition	Atmospheric Particles											
Bacteria	Fungi											
Pyrolysis-MS/MS												
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A series of experiments are reported which examined the susceptibility of the pyrolysis-mass spectrometry (Py-MS) method in identifying bacteria which had been subjected to changes in growing time, killing method and diversity in the data set. In spite of these complications, Py-MS with pattern recognition techniques was successful where the number of bacteria in the data set was less than 16. The results from a large data set, containing 47 bacteria indicated that the number of samples exceeded the capabilities of the pattern recognition technique. - Continued												

DTIC
ELECTE
FEB 05 1987

S **D**
E

AD-A176 672

DTIC FILE COPY

20. continued - Abstract

Experiments using Py-MS/MS showed that this technique could provide enough selectivity to overcome the pattern recognition problems in large sample groups. Analyses were conducted on mixtures containing bacteria in a background which might be found on a battlefield. The complexity of the mixture and the level of the bacteria seriously affected the ability to identify the bacteria. Py-MS/MS with pattern recognition techniques will probably be necessary for future development of the technique.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

The Feasibility of Using Pyrolysis-Mass Spectrometry
and Pyrolysis-MS/MS with Pattern Recognition for
the Identification of Biological Materials

Final Report

Kent J. Voorhees
Steven L. Durfee

January 7, 1987

U.S. Army Research Office
Contract #DAAG29-85-K-0199

to

Colorado School of Mines
Golden, CO 80401

Approved for Public Release;
Distribution Unlimited

TABLE OF CONTENTS

EXECUTIVE SUMMARY

A. RESEARCH BACKGROUND AND GOALS.....	3
B. INTRODUCTION.....	4
C. EXPERIMENTAL.....	5
1. Pyrolysis Method.....	5
2. Data Analysis.....	7
a. Pretreatment.....	7
b. Components Analysis.....	8
c. Graphical Rotation.....	8
D. RESULTS	9
1. Effect of Colony Age.....	9
2. Effect of Killing Method.....	16
3. Classification of a Large Data Set.....	20
4. Triple Quadrupole Techniques.....	31
5. Mixture Analysis.....	41
a. Simple Mixtures of Moderate Concentration.....	41
b. Mixtures of Lower Concentration.....	47
E. CONCLUSIONS.....	49
F. REFERENCES.....	54

EXECUTIVE SUMMARY

The goals of research outlined in the original proposal to the U.S. Army Research Office (ARO) have been accomplished during the first year of the program. At the start of the study, the identification of bacteria was limited to specimens collected and cultured under highly reproducible conditions.

As a direct result of the reported work, the understanding of the merits and limitations of single stage pyrolysis-mass spectrometry (Py-MS) for the rapid identification of target bacteria and molds in the presence of varying amounts and types of atmospheric background has increased. Also, the applicability of triple quadrupole MS/MS has been assessed.

Experiments in which a set of bacteria was grown to different stages of growth showed that the age of the culture had little or no effect on the ability of a linear classifier (using Py-MS data) to model the bacteria within the range of viability. In an experiment to investigate the effect of different methods of killing, all of the bacteria were correctly identified without regard to the method (if any) used to kill them. Simple mixtures of bacteria in this experiment were correctly identified 75 percent of the time using a K-nearest neighbors approach with only two false positives.

A large Py-MS data set consisting of 28 species of bacteria, 2 yeasts and 2 fungi, with several different strains for some of the species, was examined. The individual species of bacteria could be discriminated for about a third of these bacteria, indicating that the uniqueness of the bacterium and the complexity of the data space modeled affect the efficiency of a classification technique.

Pyrolysis-triple quadrupole MS experiments were performed using parent ion and daughter ion modes. The parent ion mode proved most useful for identification. Characteristic parent-daughter ion pairs were found for all of the bacteria tested. A methodology was developed for isolating such pairs for other bacteria. Since the parent-daughter pairs were unique for the bacteria tested, this appears to be a very rapid, accurate and potentially sensitive method for detecting bacteria from mixtures.

Pyrolysis-MS mixture experiments were performed for a set of four species of bacteria in the presence of varying contributions from two potential contaminants. The presence of bacteria was readily detected, and, once found, the species of bacteria present could be determined. Mixture experiments involving a more complex background (seven completely different contaminants with lower concentrations of seven different species of bacteria) could not be classified reliably. These results strongly suggest that the pyrolysis-triple-quadrupole approach is necessary for identifying bacteria in such mixtures.

A. RESEARCH BACKGROUND AND GOALS

A proposal entitled "The Feasibility of Using Pyrolysis-Mass Spectrometry and Pyrolysis/MS/MS with Pattern Recognition for the Identification of Biological Materials" was funded by ARO to the Colorado School of Mines. The period of the proposal was August 1, 1985 to September 30, 1986. The following were summarized in the original proposal as the goals for the proposed work:

- 1) determine the possibility of using characteristic mass spectral ion data from the pyrolysate of pure biomaterials for subsequent identification of the organism;
- 2) apply present state-of-the-art Py-MS mixture analysis techniques to simulated battlefield aerosols containing biological materials;
- 3) assess the feasibility and need of using tandem mass spectrometry with characteristic ion data for identification of aerosol mixtures.

The original proposal suggested that fungi, bacteria and viruses would be studied. Because of the lack of suitable virus samples, only bacteria and fungi have been pyrolyzed. The remainder of this report addresses the research conducted using a diverse group of bacteria and fungi.

B. INTRODUCTION

Most of the work to date involving pyrolysis of microorganisms has focused on specific biopolymers (1-4) or whole (intact or lyophilized) bacteria (5-7) using pyrolysis gas chromatography (Py-GC) or pyrolysis gas chromatography/mass spectrometry (Py-GC/MS). This work has been capably reviewed recently by Bayer and Morgan (8), and earlier by Meuzelaar et al. (9) and Irwin (10). Gas chromatographic techniques were deemed unsuitable for the present project because of the length of time per analysis, sensitivity considerations, and sample preparation. In addition, gas chromatograms are difficult to digitize reliably, which introduces an additional difficulty for computerized pattern recognition techniques.

Risby and Yergey (11) were among the first to pyrolyze whole organisms directly into a mass spectrometer. Their approach involved linear programmed thermal degradation, which gave a time profile of volatiles generated from ambient to 400°C. Using a simple binary decision tree approach, ten species of bacteria were correctly classified. Again, the method was time-consuming and the results were not easily digitized.

The Curie-point Py-MS/pattern recognition method has been applied successfully to the classification of closely-related bacteria such as several Bacillus species (12), Streptococcus species (13) and Legionella species (14). This approach is rapid and allows for reliable digitization of the data. In addition,

the procedure has been applied to numerous other materials requiring analysis of high molecular weight mixtures.

Another unusual Py-MS technique is the analysis of single particles. The Jet Propulsion Laboratory (15) has designed a unique inlet system connected to a mass spectrometer to detect single biological particles. Microorganisms in air are introduced directly into the MS in the form of a particle beam where they are individually pyrolyzed, ionized and analyzed with an electron optical ion detector (EOID).

The present study differs significantly from clinical applications where samples are cultured under controlled laboratory conditions in pure form, for reproducible lengths of time, on consistent, sometimes selective media, and are treated in the same fashion during the analysis. Furthermore, the clinical techniques usually target one or two organisms out of a limited number of possible bacteria. This makes the mathematical method of identification much simpler because of a previous pretreatment, selective media, or because the sample has been collected in an environment such as specific bodily fluids which support a limited number of possible microorganisms.

C. EXPERIMENTAL

1. Pyrolysis Method

Curie-point pyrolysis MS was used for this study. Sample preparation for Curie-point pyrolysis was straightforward.

Bacteria were grown on general-purpose media, such as trypticase soy agar, in petri dishes. A portion of the surface showing a good growth of bacteria was gently scraped with a clean 510°C Curie-point (ferromagnetic) wire. Alternatively, for samples in suspension, five microliter aliquots of the suspension were applied to a rotating ferromagnetic wire using a syringe (9). The wire, anchored in a cylindrical pyrex tube, was introduced into the high-vacuum pyrolysis inlet of the mass spectrometer on a probe. After positioning the sample axial to a high frequency coil, a radio-frequency current was applied to the coil, which resulted in the wire rising to its Curie-point temperature.

The pyrolyzate then was passed into an expansion chamber, from which it was slowly introduced into the electron ionization source of the quadrupole mass spectrometer. Low energy, 15 eV electrons were used to minimize the amount of fragmentation due to ionization. At this point the mass spectrometer was scanned repetitively. Fifty scans in the mass range 10 to 300 amu at a scan rate of 1500 amu/s. were summed and stored as a single spectrum. Details of the mass spectrometer and analysis techniques are reported in ref. 9.

Tandem MS experiments were conducted using a Curie-point gas chromatography inlet connected directly to an Extrel triple quadrupole MS. The pyrolysis inlet (similar to the one described in ref. 16) was connected through a 4 meter length of 250 μ m diameter methyl silicone capillary column connected directly to the chemical ionization (CI) source. The pyrolyzer/capillary interface was held at 300°C, the GC oven at 230°, the transfer

line at 250°, and the chemical ionization source at 200°C. Methane was used for the carrier gas and the reagent gas. There was no evidence of chromatographic separation in the column; the length was chosen to control the flow rate into the MS at about 2 mL/min while maintaining a significant positive pressure in the pyrolysis inlet. The resulting pressure near the CI source was 2×10^{-4} torr. Argon was introduced as a collision gas at a pressure which produced the greatest abundance of daughter ions. "Daughter spectra" were collected and stored as individual scans in a mass range of 10-240 amu. The scans for each pyrolysis were averaged. Three pyrolysis replicates of each bacterium were generated for a total of nine spectra. A 510°C Curie-point wire was used for all pyrolyses.

2. Data Analysis

a. Pretreatment

The pretreatment of the pyrolysis mass spectra involved normalizing the data by dividing the peak intensities in the mass spectrum by the summed intensity of the peaks in the central part of the spectrum. This reduced the effects of background and sample loading (17). Spectra were then merged into one large data matrix, with the rows being samples and the columns being mass units.

The question of whether to autoscale (18) is not an obvious one. Experience in this laboratory has shown that a variety of classification techniques provide better results after autoscaling. In order to assess the importance of autoscaling to the classification of bacteria, both unautoscaled and autoscaled approaches were used in each of the described studies.

b. Component analysis

An ordered set of numbers can be thought of as representing a point in an n-dimensional space, such as the familiar use of two-dimensional Cartesian coordinates to display equations or ordered pairs of two variables. This familiar two-dimensional display is useful because the distance between points is a measure of the similarity between the objects they represent. The same is true of ordered multiplets of data such as mass spectra, although the entire space may not be visualized accurately on a flat surface. However, the points in a higher-dimensional space may be accurately projected into two-dimensional coordinates by using principal components analysis (19). This approach provides the minimum number of independent projections necessary to completely reproduce the information in data with a large number of measurements within experimental limits of error. The ARTHUR software package (19) was used for all principal components analyses.

c. Graphical rotation

The principal component projections may be rotated two at a time in order to search the space for suspected structure (20). Each of the orthogonal axes is referred to as a factor, the value of each factor for a particular sample is referred to as a score, and the projection of the two factors is called a Karhunen-Loeve (KL) projection. As with conventional Cartesian coordinates, the distance between points is a measure of their similarity, with respect to the factor axes represented.

It is possible to obtain information about the portions of the mass spectrum responsible for the separation of data points in the projected factors. The loadings are the correlation coefficients between the factors and the peaks in the mass spectrum as well as the cosines of the angles between the mass axes and the principal components which define the data space. The loadings are contained in the eigenvectors of the correlation or covariance matrix. The corresponding eigenvalues are proportional to the amount of variance contained in the factor. If the loadings are large and positive, a given mass is correlated with the positive features on the respective KL plot. At any point following rotation, the original data set may be reproduced within experimental error by taking the matrix product of the loadings and the corresponding scores. By plotting the loadings on a bar plot similar to a conventional mass spectrum, the peaks in the mass spectrum responsible for apparent trends in the KL projection may be observed.

D. RESULTS

1. Effect of Colony Age

The successful application of Py-MS for the identification of a diverse set of bacteria demands that spectral differences caused by the stage of growth be less than the spectral differences between organisms. In order to verify this point, the six species of bacteria listed in Table 1 were cultured to different stages of growth. The cultures listed in this table were

obtained from the Colorado School of Mines collection. Each of the bacteria was grown under two conditions as listed in Table 1. Six of the bacteria grew too slowly to obtain specimens at short incubation times. For these, only the 48 hr. sample was used. All of the bacteria were grown in petri dishes on trypticase soy agar at 28°C. The bacteria were pyrolyzed according to a randomized complete block design (21).

Table 1. Number of samples of bacteria in different stages of growth.

<u>Species</u>	<u>24-hour</u>	<u>48-hour</u>
<u>Proteus vulgaris</u>	(0)	(3)
<u>Micrococcus luteus</u>	(0)	(3)
<u>Pseudomonas fluorescens</u>	(3)	(3)
<u>Escherichia coli</u>	(3)	(3)
<u>Staphylococcus aureus</u>	(0)	(3)
<u>Bacillus subtilis</u>	(2)	(3)
<u>Enterobacter aerogenes</u>	(3)	(3)
<u>Providencia alcalifaciens</u>	(3)	(3)
<u>Streptococcus faecalis</u>	(0)	(3)
<u>Streptococcus salivarius</u>	(0)	(3)
<u>Bacillus cereus</u>	(12)	(12)
<u>Bacillus polymyxa</u>	(0)	(3)

Figure 1 shows sample Py-MS spectra for E. coli. In the spectrum, mass 117 was dominant. Figure 2 is a KL projection (following rotation) which shows that three of the four strains

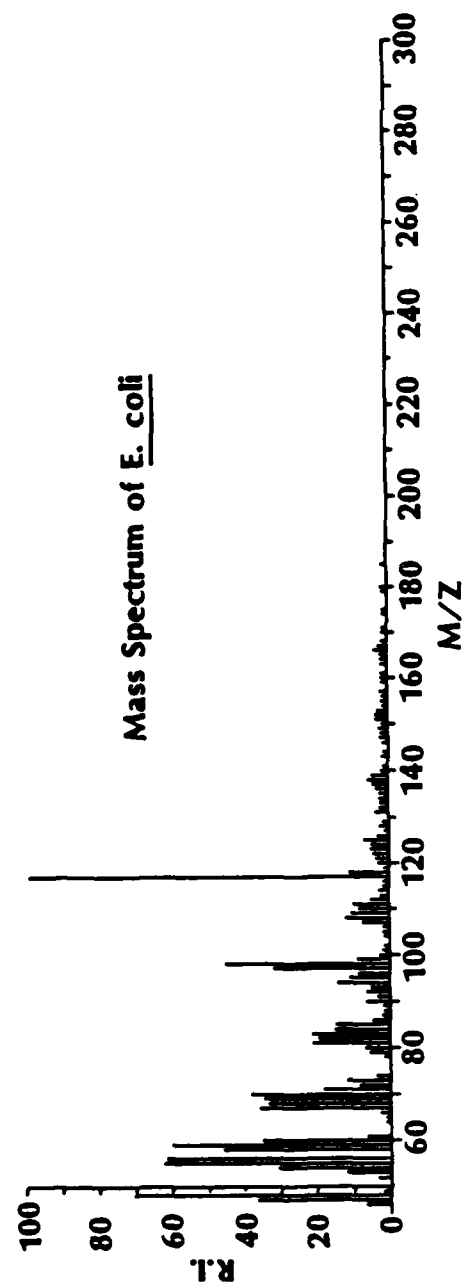


Figure 1. Py-Mass spectrum of E. coli.

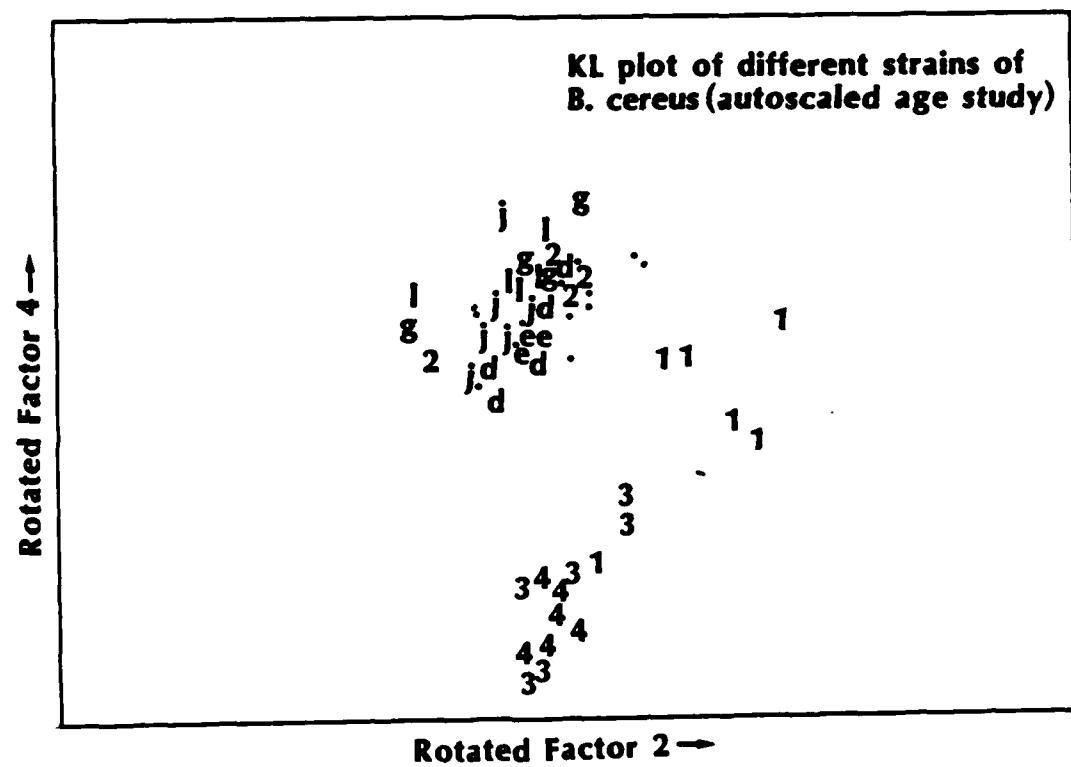


Figure 2. KL Plot showing separation of the different strains of *B. Cereus* (1-4).

of B. cereus (indicated by the characters '1', '3', and '4') were separated from the other strain ('2') and the rest of the bacteria. Also, strains '3' and '4' appear to be similar, but are mostly separated from strain '1'.

Figure 3 shows another KL projection (following rotation) in which E. coli (e), B. subtilis (g) and P. alcalifaciens (l) were readily separated from the other species of bacteria. Any line which can be constructed on these plots separating one species from another is essentially a discriminant function. Such a function, when used with a training set can be utilized to identify unknown bacteria. Clearly discriminant functions exist which separate e, g, and l compared to all the other bacteria. KL plots (not shown) were generated such that all the bacteria in this data set were discriminated from each other despite the age differences. Because of the separations observed, factor spectra were generated showing the peaks responsible for the separations; the one for P. fluorescens is shown in Figure 4. Positive peaks in the factor spectrum are peaks which are consistently more abundant in the spectrum of P. fluorescens when compared to the other bacteria spectra in the set. The indicated masses invite speculation, but the identity of the peaks cannot be determined from the single stage mass spectra alone because of the enormous chemical complexity of biological systems and the fact that each of the peaks in a pyrolysis mass spectrum probably includes multiple isomers.

Clearly the age of the bacterial colonies does not adversely affect the classification. By focusing on the dichotomy of old

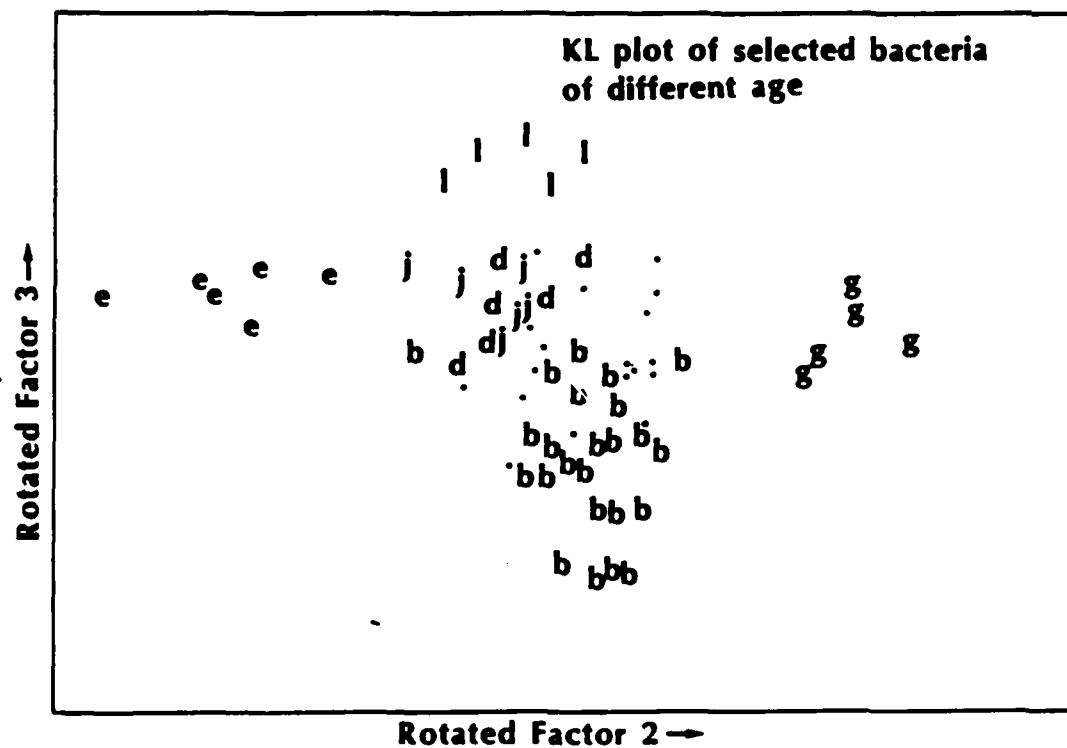


Figure 3. KL Plot of showing separation of E. Coli (e), B. cereus (b), B. subtilis (g) and P. alcalifaciens (l).

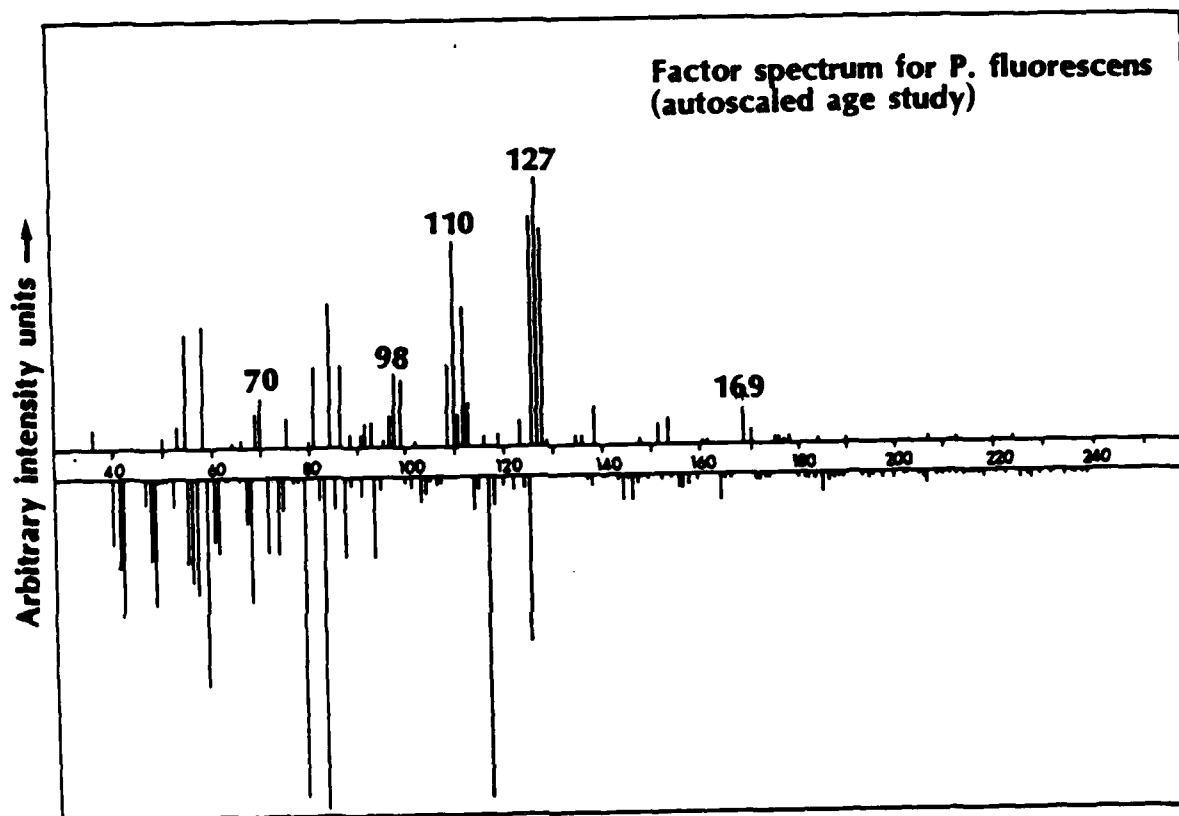


Figure 4. Factor spectrum of *P. fluorescens* (Top Spectrum).

versus young, no non-species specific factor was found which could uniquely separate the two ages in the data. If there are specific features which indicate age within species, they were not apparent from the comparison of samples of 3 old to 3 young bacteria.

2. Effect of Killing Method

When killed, bacteria necessarily undergo chemical changes due to the killing process. Presumably, the specific changes include uniform changes due to cessation of metabolism. They also include outright chemical reactions with the bactericide (incorporation), loss of material due to solvation, and reactions with atmospheric oxygen. Killing due to heat probably includes air oxidation and thermolysis. In a clinical study, a single killing technique, usually the treatment with ethanol, is used exclusively.

In a study of killing methods, four species of bacteria, P. vulgaris, P. fluorescens, E. coli and E. aerogenes, were each subjected to five different sample preparations (see Table 2). One preparation was direct sampling of live colonies from a trypticase soy agar. The other four sample treatments involved suspensions. These samples were originally grown on plates, then suspended in deionized water by introducing about two mL of water, followed by gently rubbing the surface of the agar with a stirring rod to suspend the bacteria. One set of samples resulted from direct analysis of these suspensions. Another resulted from autoclaving the suspension and then analyzing. The final two sample sets were suspensions in 7:8 solutions of methanol:water and ethanol:water.

Table 2. Bacteria and treatments in killing experiment.

I. Identification:	<u>Code</u>
A. <u>P. vulgaris</u>	a
B. <u>P. fluorescens</u>	d
C. <u>E. coli</u>	e
D. <u>E. aerogenes</u>	j
II. Treatment:	
A. Suspensions	
1. No treatment	
2. Autoclaved 15 min.	
3. Methanol	
4. Ethanol	
B. Plates	
III. Replicates: (3)	

Figure 5 is a KL plot showing the separation for pyrolysis data of bacteria P. fluorescens (d) and E. aerogenes (j). Similar plots (not shown) were obtained for the other bacteria. In addition to the single samples in Table 2, ten mixtures were prepared as follows: a+d, d+e, d+j, e+j, a+d+e, a+d+j, a+e+j and d+e+j. These are indicated as 'm' on Figure 5. With the exception of two poorly resolved spectra of e, all of the bacteria could be correctly identified despite differences in killing.

K nearest-neighbors (KNN) analysis (19) was used as a simple test of clustering. Table 3 shows a 94 percent success rate at

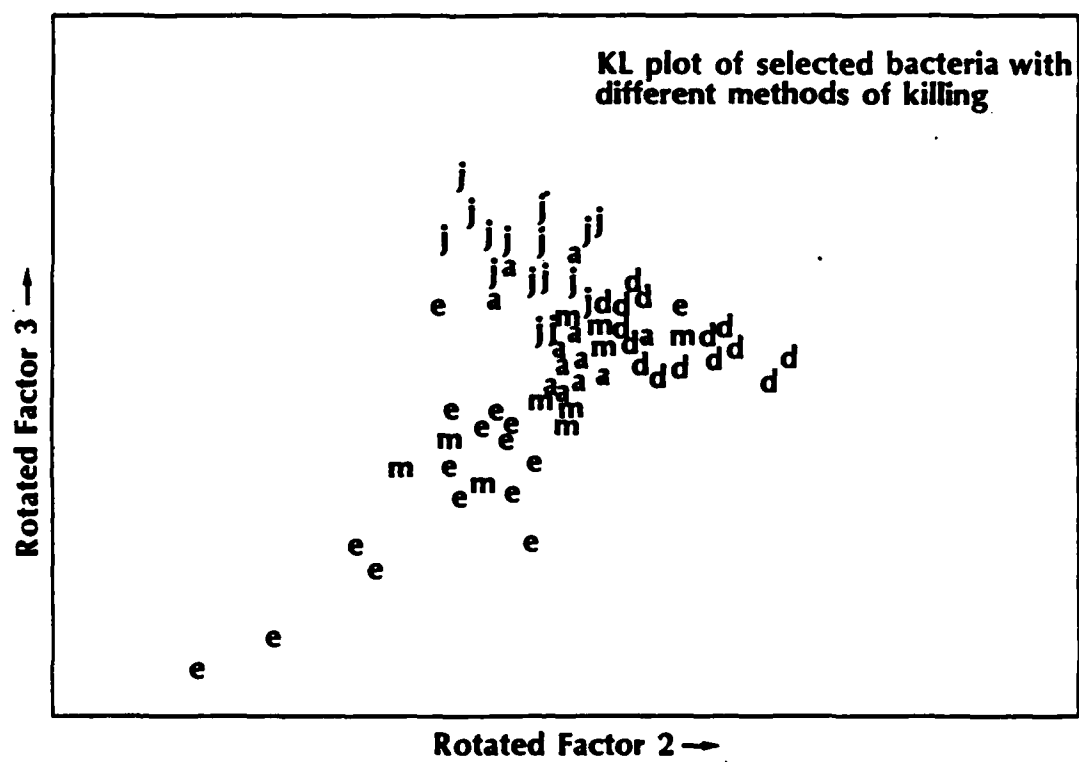


Figure 5. KL Plot of selected bacteria killed by different methods.

describing the training set (known bacteria) according to the conventional KNN cluster analysis approach. The mixtures were also classified as containing whatever components were represented among the ten nearest neighbors (Table 4). Using this very simple approach, the components were identified 75 percent of the time, with less than 10 percent false positive.

Table 3. Computed class for training set based on one nearest neighbor.

<u>True</u>	<u>P. vulgaris</u>	<u>Calculated P. fluorescens</u>	<u>E. coli</u>	<u>E. aerogenes</u>
<u>P. vulgaris</u>	100.0	0.0	0.0	0.0
<u>P. fluorescens</u>	0.0	100.0	0.0	0.0
<u>E. coli</u>	0.0	5.6	88.9	5.6
<u>E. aerogenes</u>	0.0	11.1	0.0	88.9

Table 4. KNN prediction of components of mixtures based on pure component training set.

<u>Actual Composition</u>	<u>10 Nearest Neighbors</u>	<u>Percent Correct</u>	<u>False Positives</u>
1,2	1,2	100%	
1,3	3	50%	
1,4	1,2,4	100%	1
2,3	3	50%	
2,4	2,4	100%	
3,4	3	50%	
1,2,3	2,3,4	100%	
1,2,4	1,2	67%	
1,3,4	2,3,4	67%	1
2,3,4	3,4	67%	

3. Classification of a Large Data Set

In order to increase the level of complexity of the data set, thirty-two different species of microorganisms, listed in Table 5, were examined. A large number of KL plots and rotations

Table 5. List of microorganisms used in diverse data set experiment. * non-bacteria. ** grown on trypticase soy agar augmented with sodium chloride.

<u>Name</u>	<u>Code</u>	<u>Replicates</u>
1. <u>Proteus vulgaris</u>	a	5
2. <u>Bacillus cereus</u> (five strains)	b	30
3. <u>Micrococcus luteus</u>	c	5
4. <u>Pseudomonas fluorescens</u>	d	5
5. <u>Escherichia coli</u>	e	5
6. <u>Staphylococcus aureus</u>	f	5
7. <u>Bacillus subtilis</u> (two strains)	g	8
8. <u>Mycobacterium phlei</u>	h	5
9. <u>Bacillus licheniformis</u>	i	5
10. <u>Enterobacter aerogenes</u>	j	5
11. <u>Streptococcus lactis</u>	k	5
12. <u>Providencia alcalifaciens</u>	l	5
13. <u>Streptococcus faecalis</u>	m	5
14. <u>Streptococcus salivarius</u>	n	5
15. <u>Bacillus polymyxa</u>	o	6
*16. <u>Candida pseudotropicalis</u>	r	5
*17. <u>Trichoderma reesei</u>	s	5
*18. <u>Saccharomyces cerevisiae</u> (two sources)	u	10
*19. <u>Penicillium notatum</u>	v	5
20. <u>Corynebacterium diphtheriae</u>	z	6
21. <u>Neisseria catarrhalis</u>	1	5
22. <u>Alcaligenes faecalis</u>	2	5
23. <u>Erwinia carotovora</u>	3	5
24. <u>Rhizobium leguminosarum</u>	4	5
25. <u>Nocardia asteroides</u>	5	5
26. <u>Bordatella sp.</u>	6	5
27. <u>Serratia marcescens</u>	7	5
**28. <u>Halobacterium sp.</u>	8	4
29. <u>Bacillus anthracis</u>	9	6
30. <u>Yersinia pestis</u>	@	5
31. <u>Francisella tularensis</u>	#	5
32. <u>Campylobacter sp.</u> (six species)	p,q,t,w,x,y	46

were required to display the structure of the Py-MS data generated in this part of the study. Figure 6 is one such plot. Since the design was not balanced (equal numbers of each bacterium), the species with the most replicates contributed the greatest to the variance, and their discriminants dominate the first few factors. For this reason, the Campylobacter species were removed from the data set. The separation of T. reesei (s) and P. notatum (v), both non-bacteria, was a dominant property of the first factor, and the separation of B. cereus (b) dominated the second. With proper rotation, these species were clearly separated (Figure 6 and 6a) and factor spectra were obtained. Using a similar approach E. coli was also separated from the other bacteria (Figure 7). The factor spectrum for E. Coli, as well as the comparable E. Coli factor spectra from the growth and killing studies (Figure 8), demonstrates that the chemical trends revealed by graphical rotation provide consistent chemical information despite differences in the comparison set. It is clear that mass 117 is the dominant feature of the spectrum of E. coli as compared to other bacteria. The consistency of factor spectra is also shown in the comparison of the factor spectrum for P. fluorescens generated from the killing study (Figure 9) and the equivalent spectrum from the age study (Figure 4). Among the prominent peaks highlighted in both spectra are masses 69, 81, 98, 110, 112, 126, and 169. These are clearly chemical components characteristic of the pyrolysis of this bacterium.

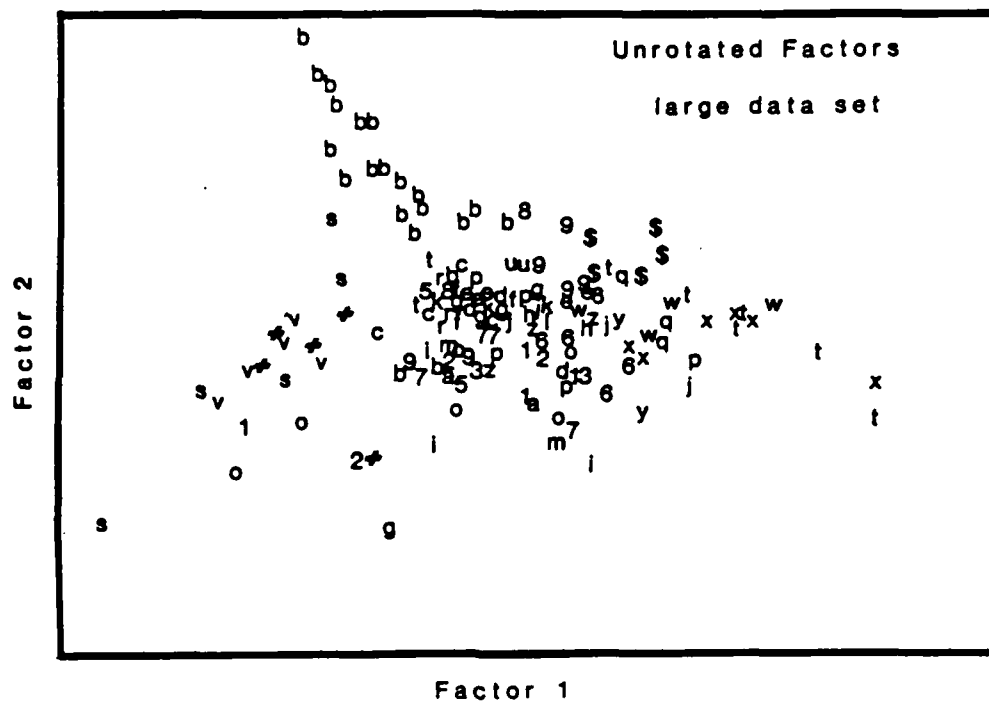


Figure 6. KL Plot of unrotated data for the Large Data Set.

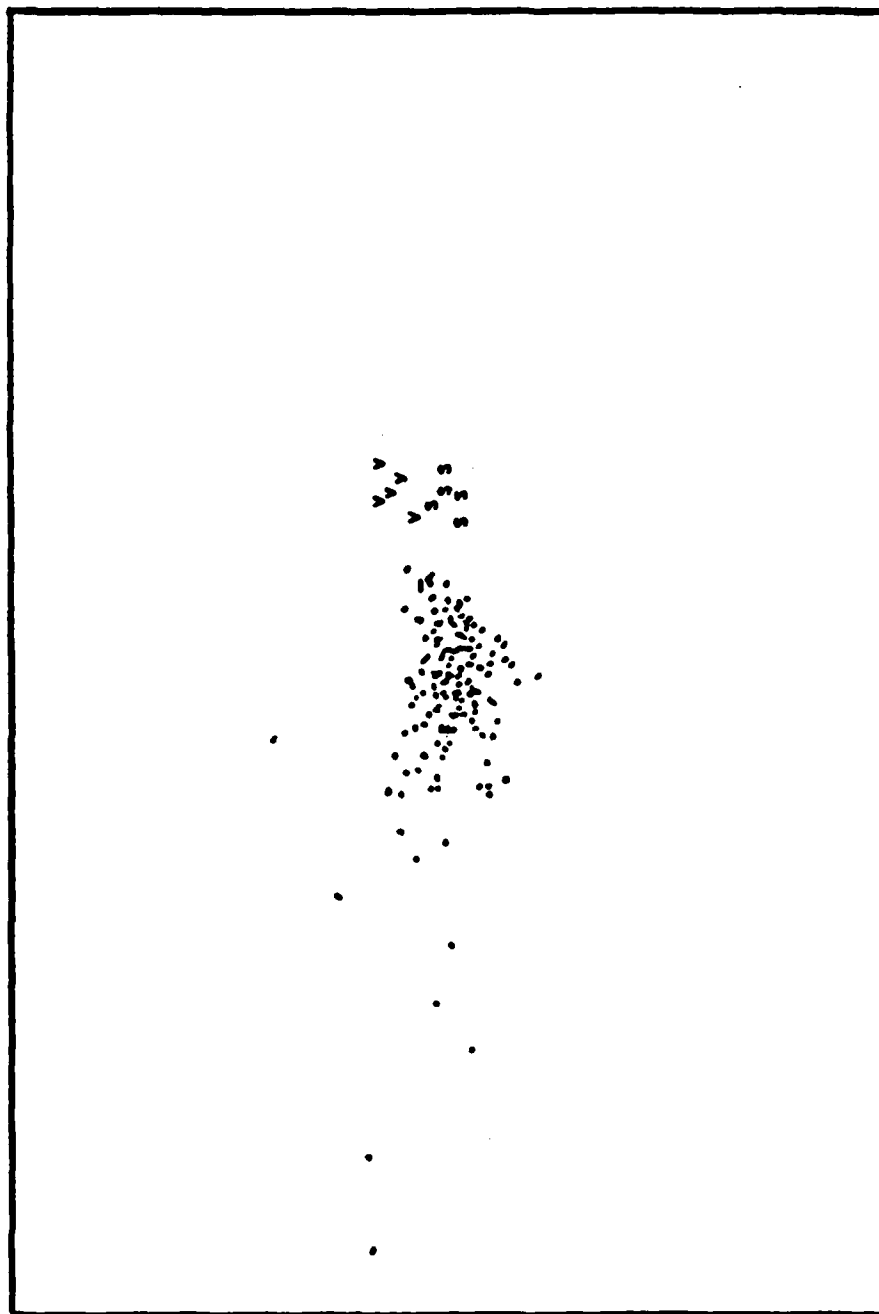


Figure 6a. KL Plot separating the two molds (v and s) from 30 different bacteria (dots).

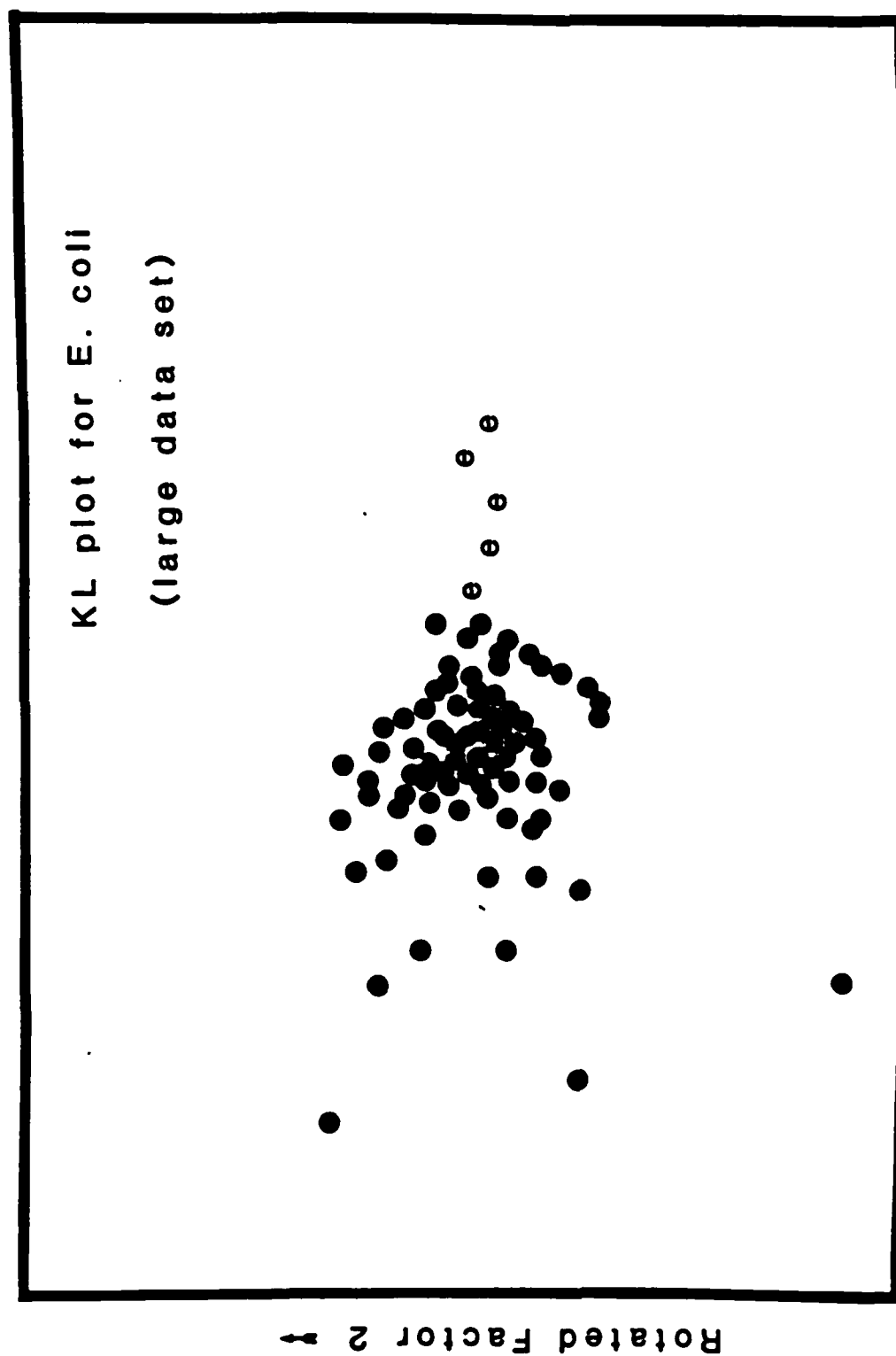


Figure 7. KL Plot showing the separation of E. Coli from the other bacteria in the Large Data Set.

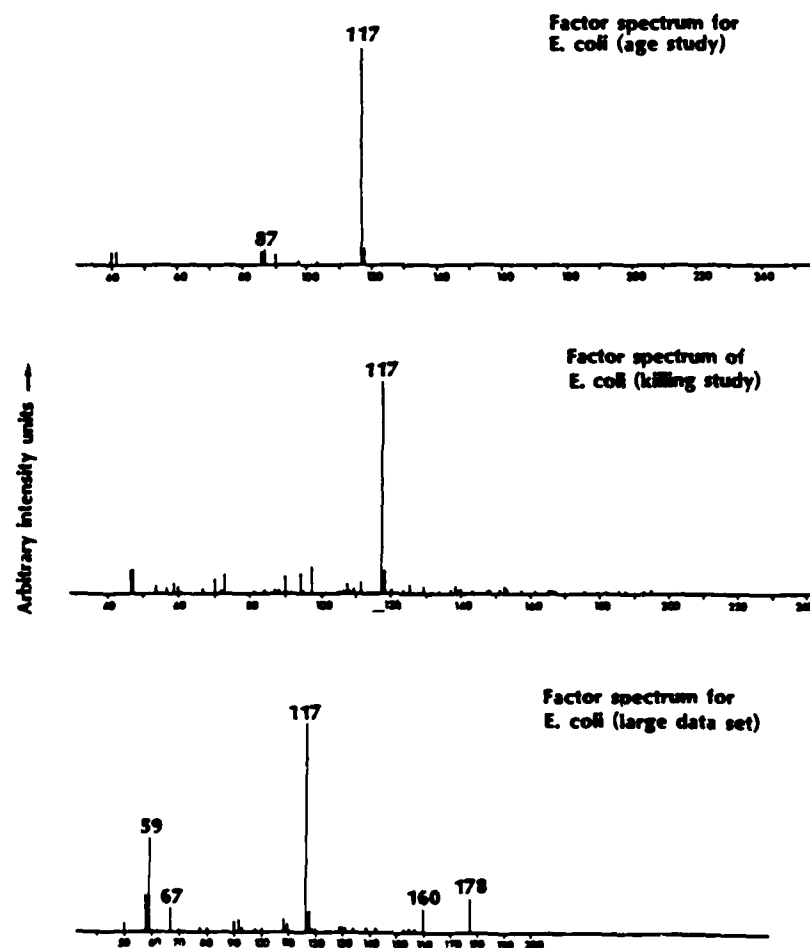


Figure 8. Factor spectra for E. Coli from three different studies.

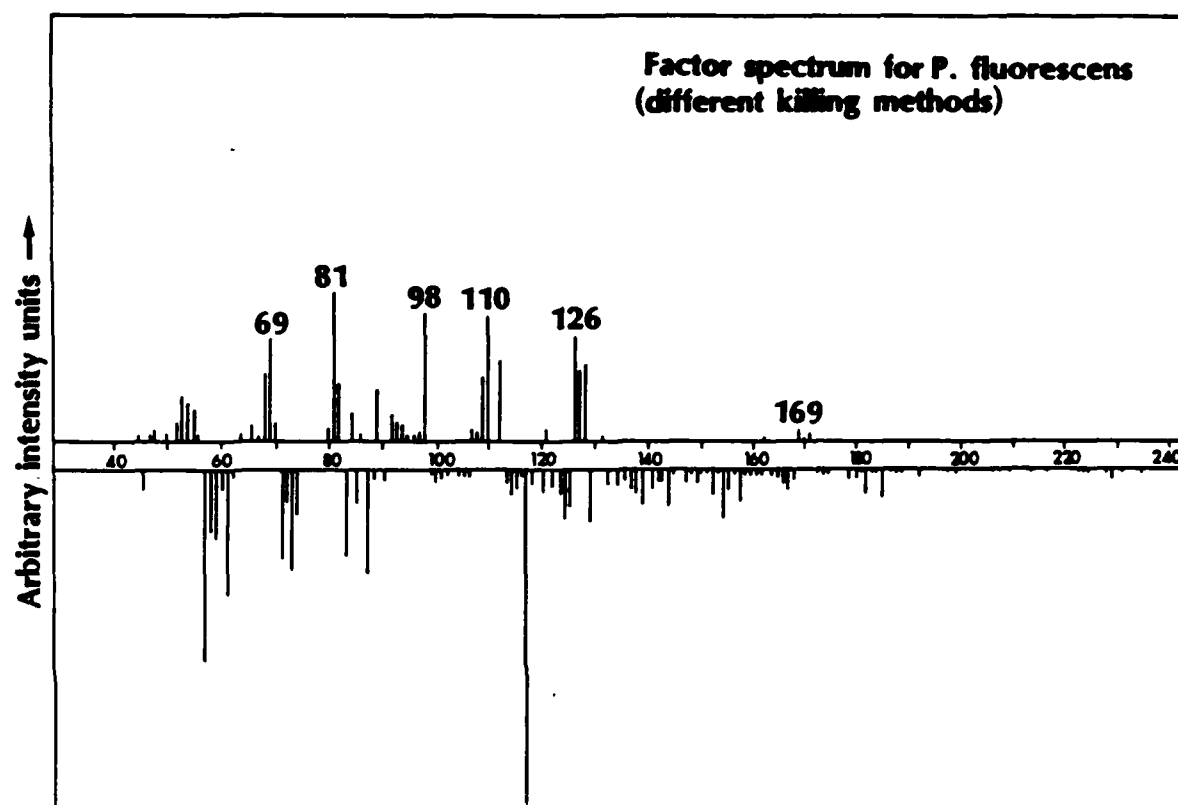


Figure 9. Factor spectrum of *P. fluorescens* from the killing study.

The chemical identity of these peaks can be speculated; however, an ancillary technique will have to be used for positive identification. For purposes of classification, the peaks listed in Table 6 were found diagnostic of the individual bacteria, including the three pathogens B. anthracis, F. tularensis and Y. pestis.

Table 6. Diagnostic Peaks for selected bacteria (EI mode).

<u>B. anthracis</u>	57, 60, 61, 94, 151
<u>B. cereus</u>	68, 86, 103
<u>B. polymyxa</u>	71, 86, 103, 117, 129, 160, 185
<u>E. coli</u>	(59), (67), 117, (160), (178)
<u>F. tularensis</u>	84, 95, 96, 103, 110, 126
<u>P. fluorescens</u>	69, 81, 98, 110, 112, 126, 127, 128, 169
<u>Y. pestis</u>	48, 94, 117
molds	58, 59, 72, 73, 85, 95, 126, 166

Comparisons between separate data sets indicate that the classification procedures for a given bacterium from one data set should apply to the same bacterium from another data set. Mathematically, this corresponds to a projection of a test bacterium onto the space modeled for the same species of bacterium in another data set. This approach has been applied

and has met with moderate success, limited only by the consistency of the instrumental parameters and the fact that classification of an unknown sample to some extent requires that the unknown has been included in the data set from which the model was drawn. Windig (22) has suggested that this is primarily a normalization effect, since negative correlations in the data (such as those indicated in the factor spectra) relate more to the several species of bacteria which the targeted bacterium is being compared to than they relate to the target itself. Thus, for an unknown to yield a negative result in a comparison, it is desirable that its spectrum contain some large peaks which are negatively correlated with the bacterium modeled. A pattern recognition technique which is not affected by the comparison set would be welcome, but no such technique with the modeling power of the component analysis approaches was found in the course of this study. Because of these arguments, test set bacteria which were not present in the training set yielded unpredictable results and generally weren't included in the projections described in this section.

Figure 10 shows a projection of the B. cereus strains from the large data set onto a model for B. cereus generated from the growth study. It includes one strain, '5', which was not available in the earlier study. With one or two exceptions, the projected strains of B. cereus were correctly modeled. Figure 11 is a similar projection of E. coli, 'e', B. subtilis, 'g', and P. alcalifaciens, 'l'. The first two bacteria modeled reasonably well, but the third was ineffectively modeled.

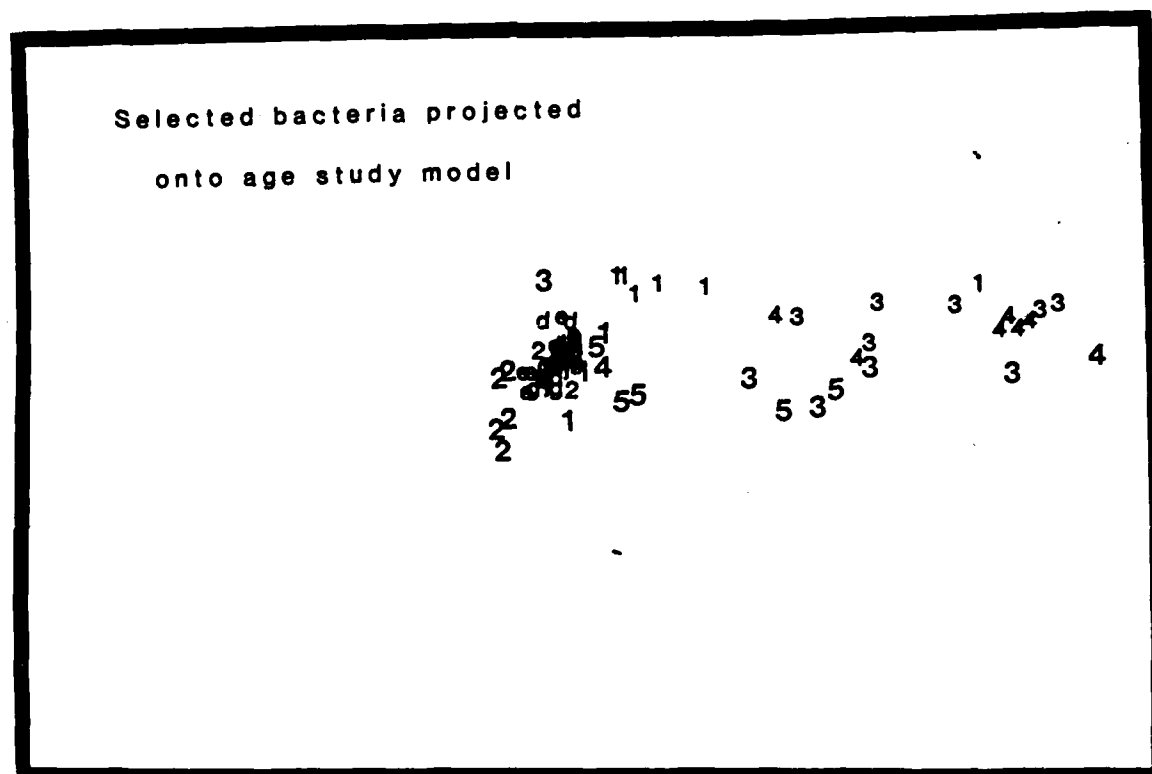


Figure 10. A model study showing B. cereus strains projected onto a KL plot where the age study data served as the model (small letters model, large letters calculated).

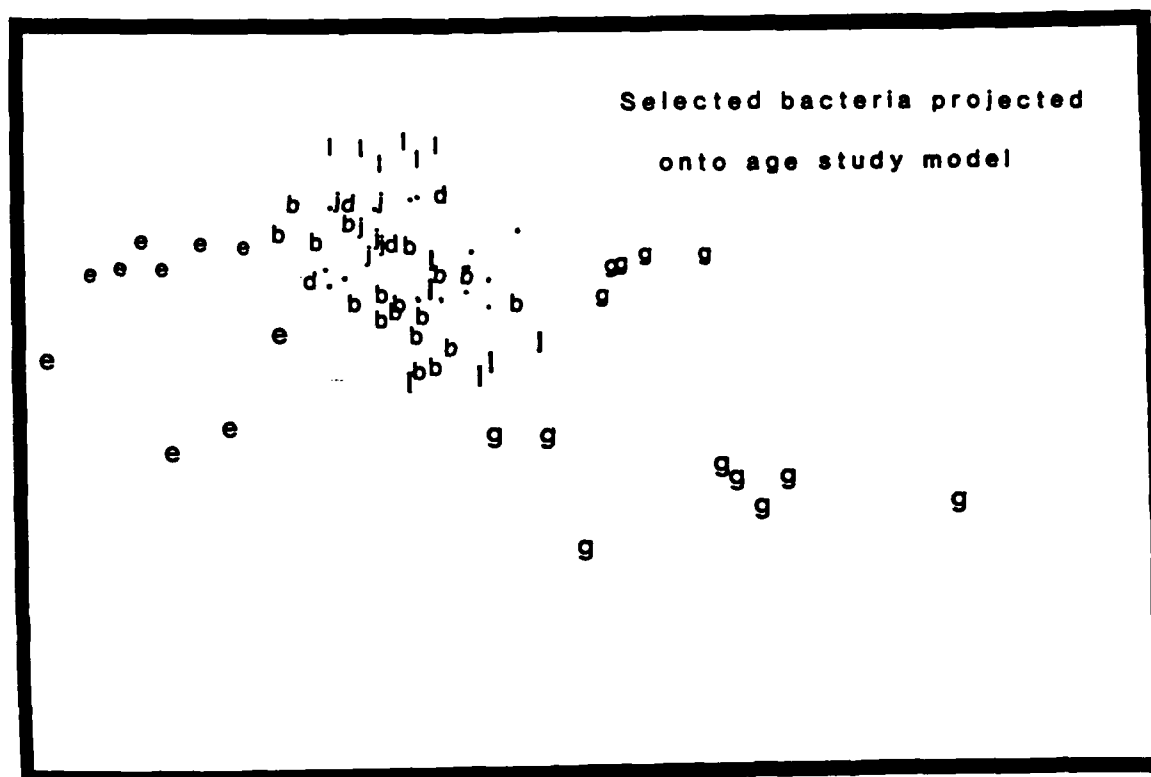


Figure 11. Additional bacteria projected onto a KL plot using the age study data as the model (small letters model, large letters projected).

4. Triple Quadrupole Techniques

Pyrolysis/tandem mass spectrometry (Py/MS/MS) permits three kinds of experiments; neutral loss scans, parent ion scans and daughter ion scans (23,24). These three scanning techniques provide a way to eliminate one of the chief limitations of conventional low resolution mass spectrometry - each peak in the spectrum of a mixture is probably not unique. An MS/MS experiment adds an extra dimension of information to the mass spectral data by allowing characterization of individual ions. The neutral loss experiment requires a judicious choice of neutral leaving groups, such as 18 amu (water) or 44 amu (CO_2). The other two scan modes each require a separate preliminary experiment to determine the parents or daughters of interest, as follows:

1. Parent ion approach: The goal of this technique is to find daughter ions which have unique parents or distributions of parents. Therefore it is first necessary to isolate the most characteristic daughter ions. One method of doing this is to hold the first quadrupole in an RF-only mode, passing all ions simultaneously, allowing the ions to collide with the collision gas, and scanning the third quadrupole to collect spectra of all the daughters. If a portion of the daughters are highly characteristic, there is a high probability that some of the characteristic daughters in the parent spectra will also be unique. Once characteristic daughters have been selected, the M^c/MS is run in the parent ion mode.

2. Daughter ion approach: This technique seeks ions which produce unique daughter spectra. In the preliminary experiment, the first quadrupole is scanned and the second and third are operated in RF-only mode, with no collision gas. This is the same as a conventional one-stage Py-MS experiment. Once characteristic ions are isolated, the MS/MS is operated in the conventional daughter ion mode, with the first quadrupole fixed at a characteristic ion. For this method to work, the molecular weights selected for the characteristic ions when fragmented must provide unique ions.

Samples of B. cereus, B. subtilis, and E. coli were pyrolyzed at 510°C in a Curie-point pyrolysis gas-chromatograph inlet. After normalization, principal components analysis was performed using the pattern recognition routines in ARTHUR to find the eight principal eigenvectors of the correlation matrix. Since none of the factors were eliminated, the component analysis was not used for its data compression properties; rather, it was used to highlight the factors containing the greatest amount of variance and to achieve the reduction from 230 correlated (oblique) masses to eight orthogonal factors.

Three (nonorthogonal) real factors were sought, one for each of the three bacteria. Figure 12 shows the best separation of B. cereus (samples labeled 'a') from B. subtilis ('b') and E. coli ('c'), so that the B. cereus spectra have positive scores on the rotated factor. Also shown on Figure 12 is the factor spectrum for the peaks in the daughter spectra on

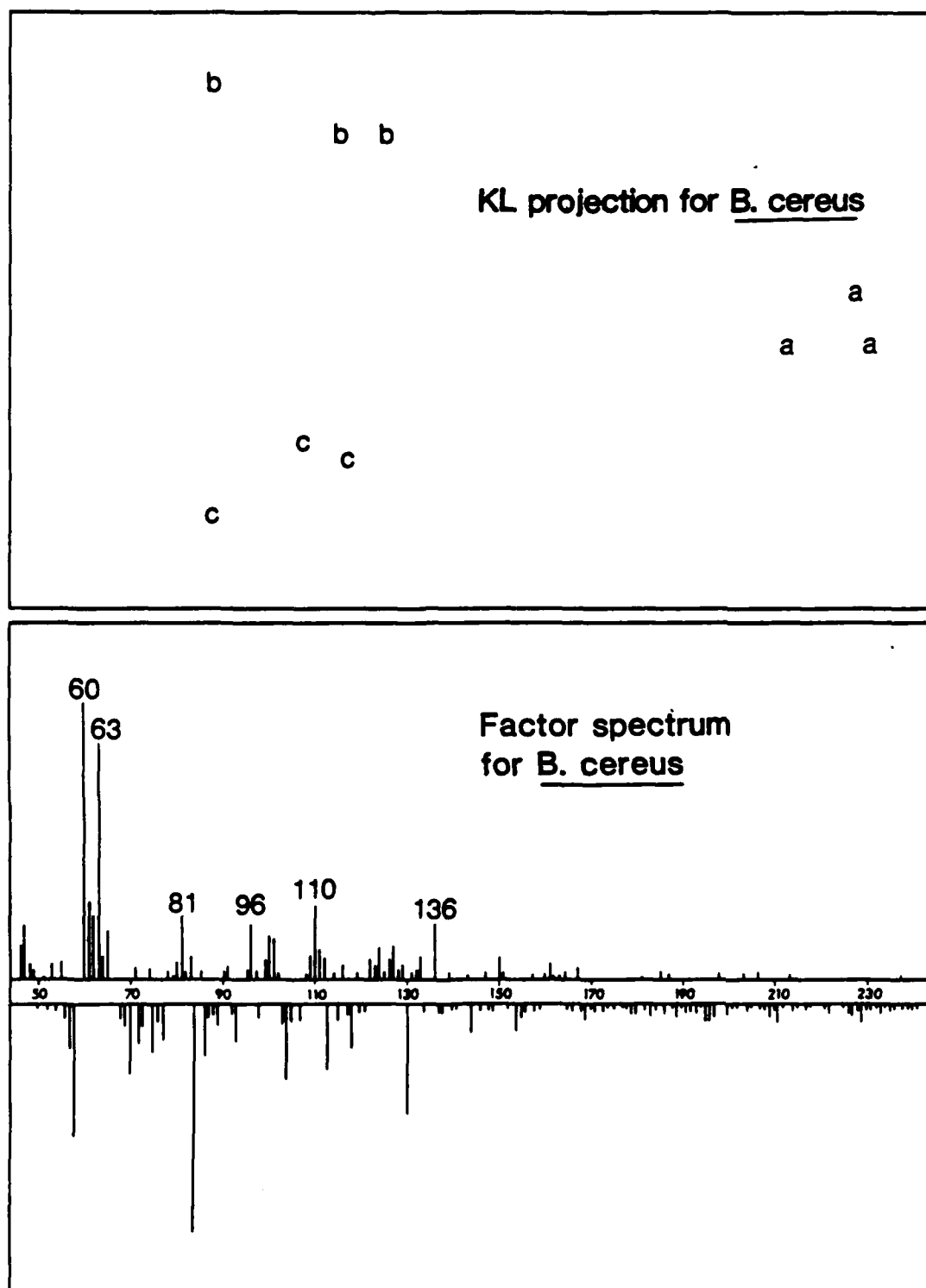


Figure 12. KL separation and factor spectrum for B. cereus Py-MS/MS data (Ignore negative spectrum).

the horizontal factor. Figures 13 and 14 show similar information for bacteria b and c. From these spectra the peaks listed in Table 7 were selected.

Not all of the peaks listed in Table 7 proved to be diagnostic. Figures 15 through 18 are successful examples of the degree of selectivity that can be gained using the triple quadrupole approach. Unique parent ions were found for all three bacteria; a result which was not possible for any of the three bacteria in single quadrupole experiments. The effect of the ionization mode on these results are unknown at this time.

Table 7. Diagnostic Daughter Peaks for Selected Bacteria (CI mode).

B. cereus	B. subtilis	E. coli
(a)	(b)	(c)
60	58	70
63	84	77
71	104	84
96	117	95
110	118	113
136	144	127
	155	130

The analysis by EI, single-stage mass spectrometry reported earlier indicated that the peaks listed in Table 6 were highly diagnostic of the listed bacteria. Similar series were found for the bacteria in the large data set. These selections should be useful for classification using MS/MS experiments. At this point, only CI MS/MS experiments are reported. The EI experiments will be done during the continuation of this project.

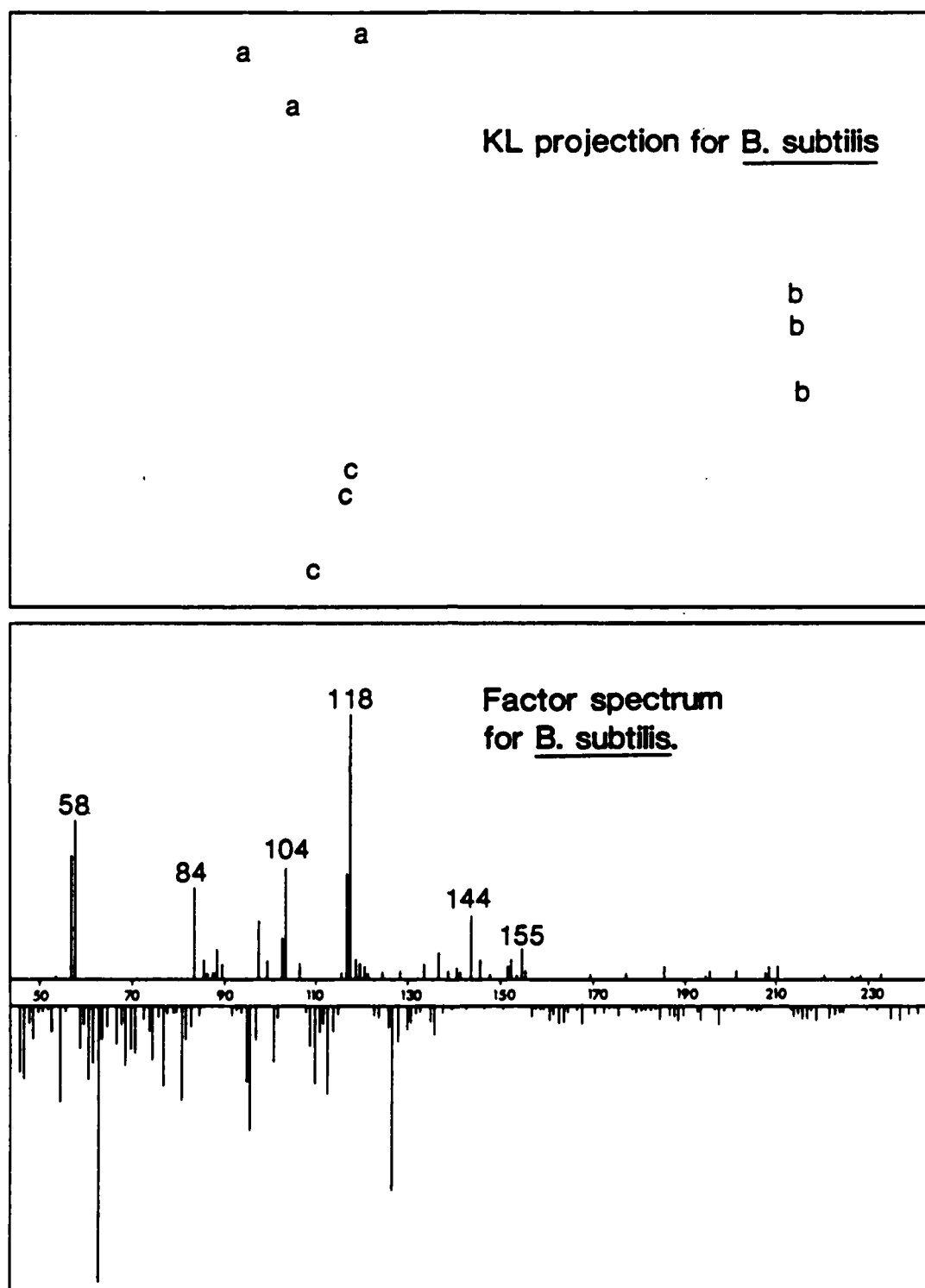


Figure 13. KL separation and factor spectrum for B. subtilis Py-MS/MS data. (ignore negative spectrum).

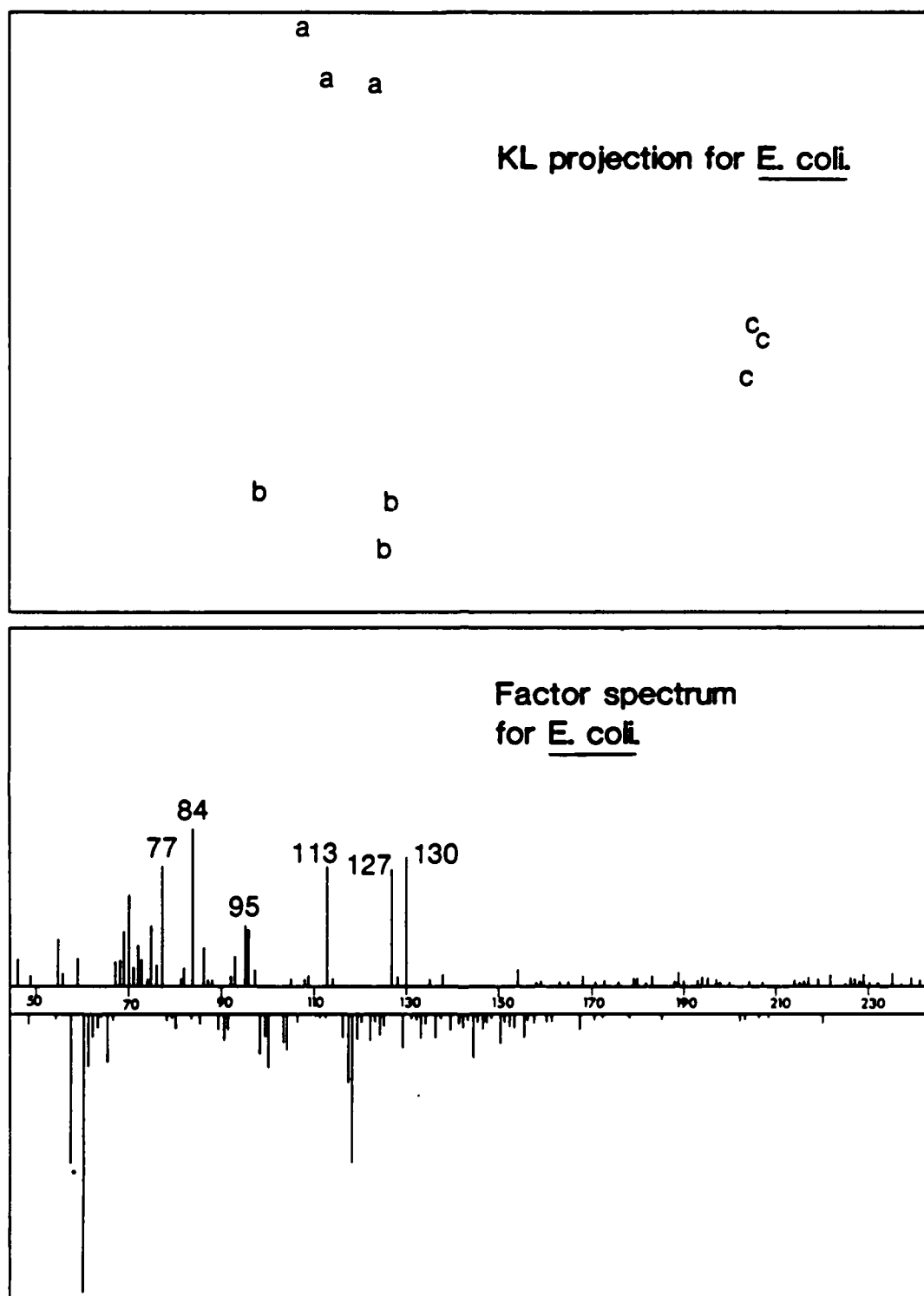


Figure 14. KL separation and factor spectrum for E. Coli Py-MS/MS data (Ignore negative spectrum).

Parents of 58

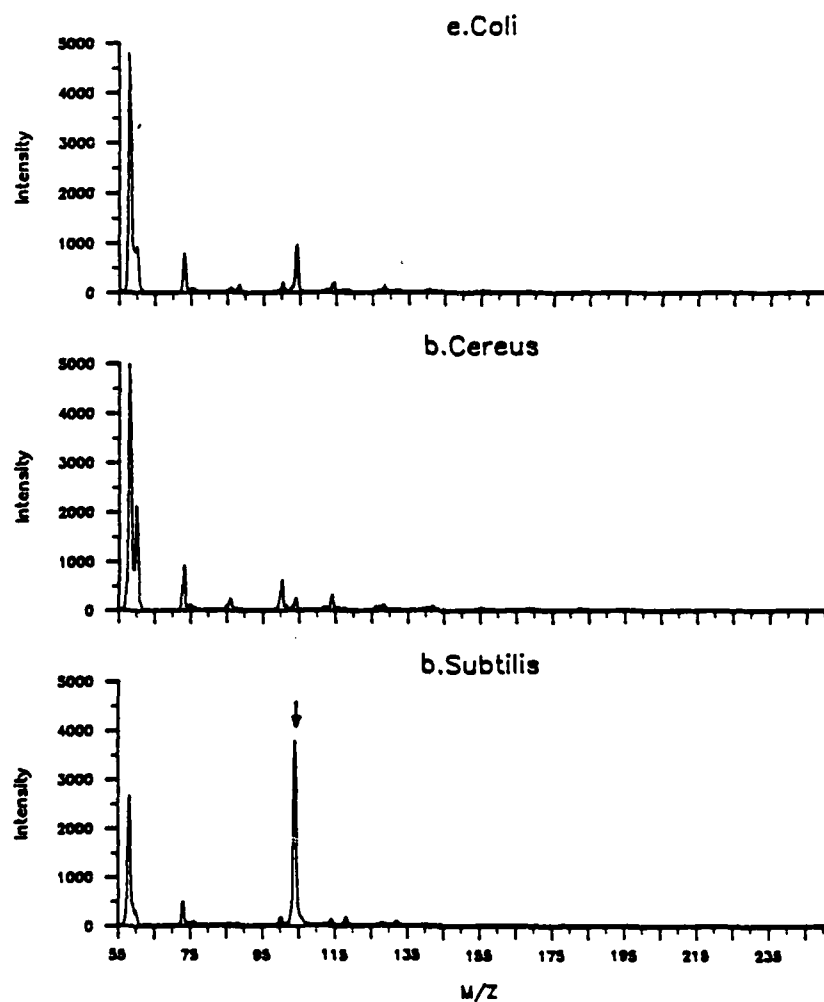


Figure 15. Parent Ion Spectra for m/z 58: Pyrolysis - MS/MS of E. coli, B. cereus and B. subtilis.

Parents of 71

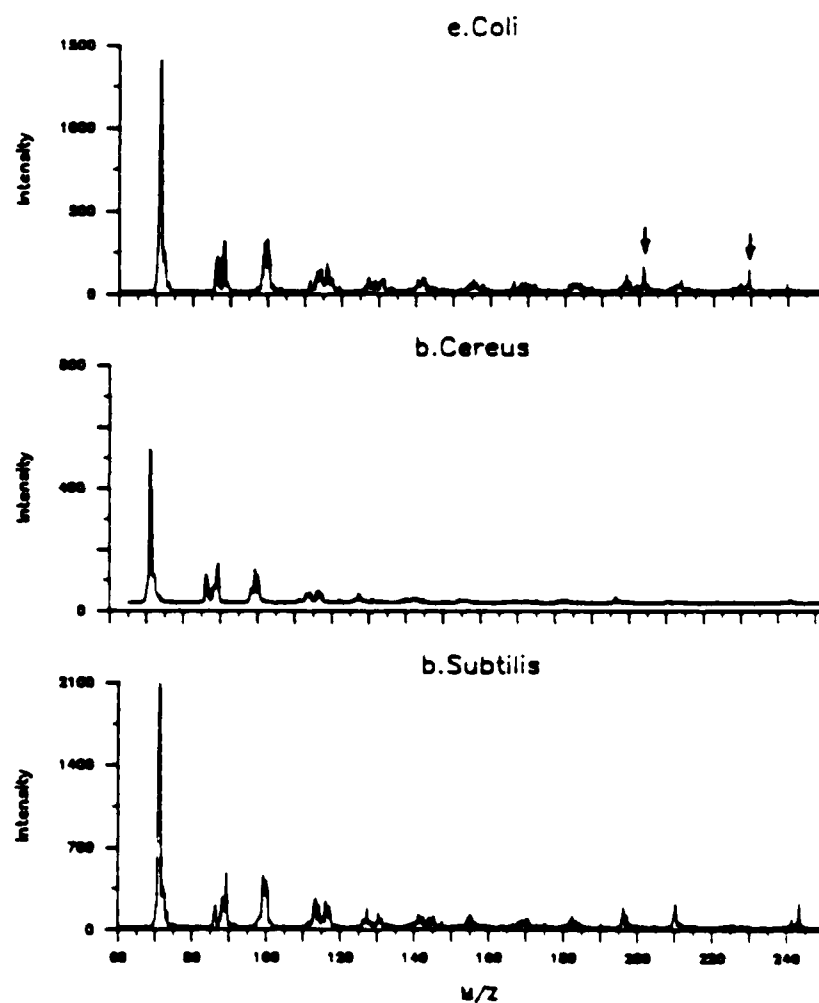


Figure 16. Parent Ion Spectra for m/z 71: Pyrolysis - MS/MS of E. coli, B. cereus, B. subtilis.

Parents of 84

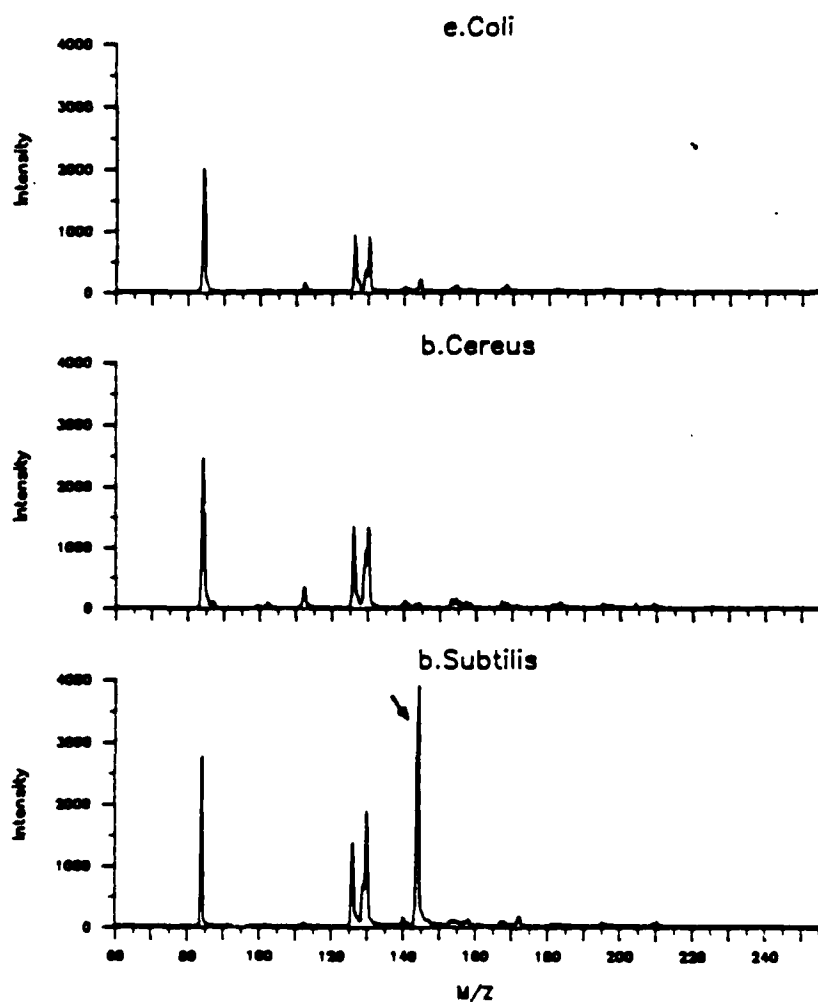


Figure 17. Parent Ion Spectra for m/z 84: Pyrolysis - MS/MS of E. coli, B. cereus, B. subtilis.

Parents of 127

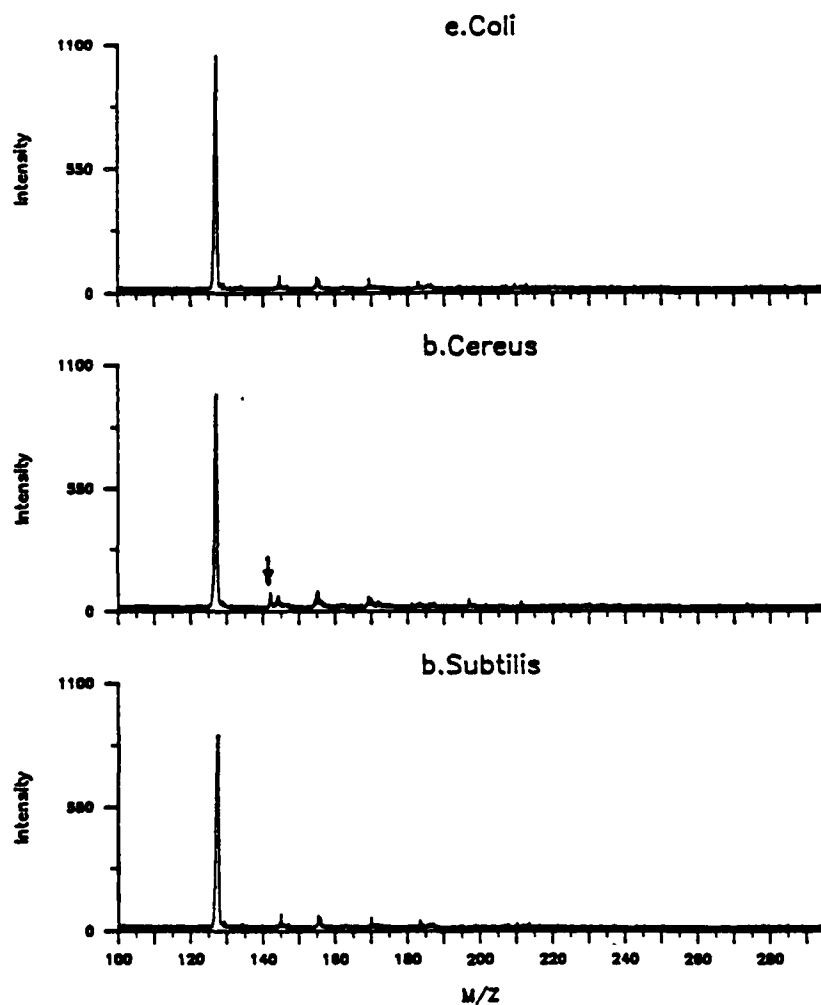


Figure 18. Parent Ion Spectra for m/z 127: Pyrolysis - MS/MS of E. coli, B. cereus, B. subtilis.

5. Mixture Analysis

a. Simple mixtures of moderate concentration

As an initial step in examining simple mixtures, forty samples containing varying amounts of pollen, particulate and each of four bacteria were formulated according to a design scheme called a simplex. The design using one bacterium is represented by the simple ternary diagram (Figure 19) and by the formulation in Table 8. Samples 1, 7 and 10 are pure bacteria, pollen and particulate. The other samples are appropriate mixtures as indicated on the diagram. The entire simplex was repeated for P. vulgaris (samples 1-10), P. fluorescens (samples 11-20), E. coli (samples 21-30) and E. aerogenes (samples 31-40). The samples were coded the same modulo 10, so that samples 1, 11, 21, and 31 each contained a pure bacterium, 2, 12, 22 and 32 each contained 2 parts bacteria with 1 part pollen, and so forth.

The pollen and particulate samples (collected on the roof of the CSM chemistry building over a two-week period using a Hi-vol sampler) were dried, weighed and suspended so that a 5 μ L aliquot was 25 μ g of solid. The amounts of bacteria were estimated visually. After the mixtures had been coated on the Curie-point wires, they were analyzed in a previously-determined random sequence.

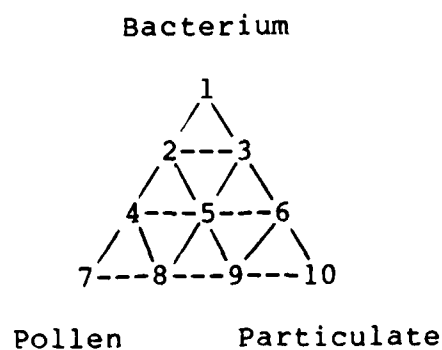


Figure 19. Ternary diagram for simplex

Table 8. Components of simplex (number of 5 μ L aliquots).

<u>Sample number</u>	<u>Bacterium</u>	<u>Pollen</u>	<u>Particulate</u>
1	3	0	0
2	2	1	0
3	2	0	1
4	1	2	0
5	1	1	1
6	1	0	2
7	0	2	0
8	0	2	1
9	0	1	2
10	0	0	3

After merging the forty spectra into one large file, the data were autoscaled and analyzed using principal components. Figure 20 is a KL plot of the first two component axes which

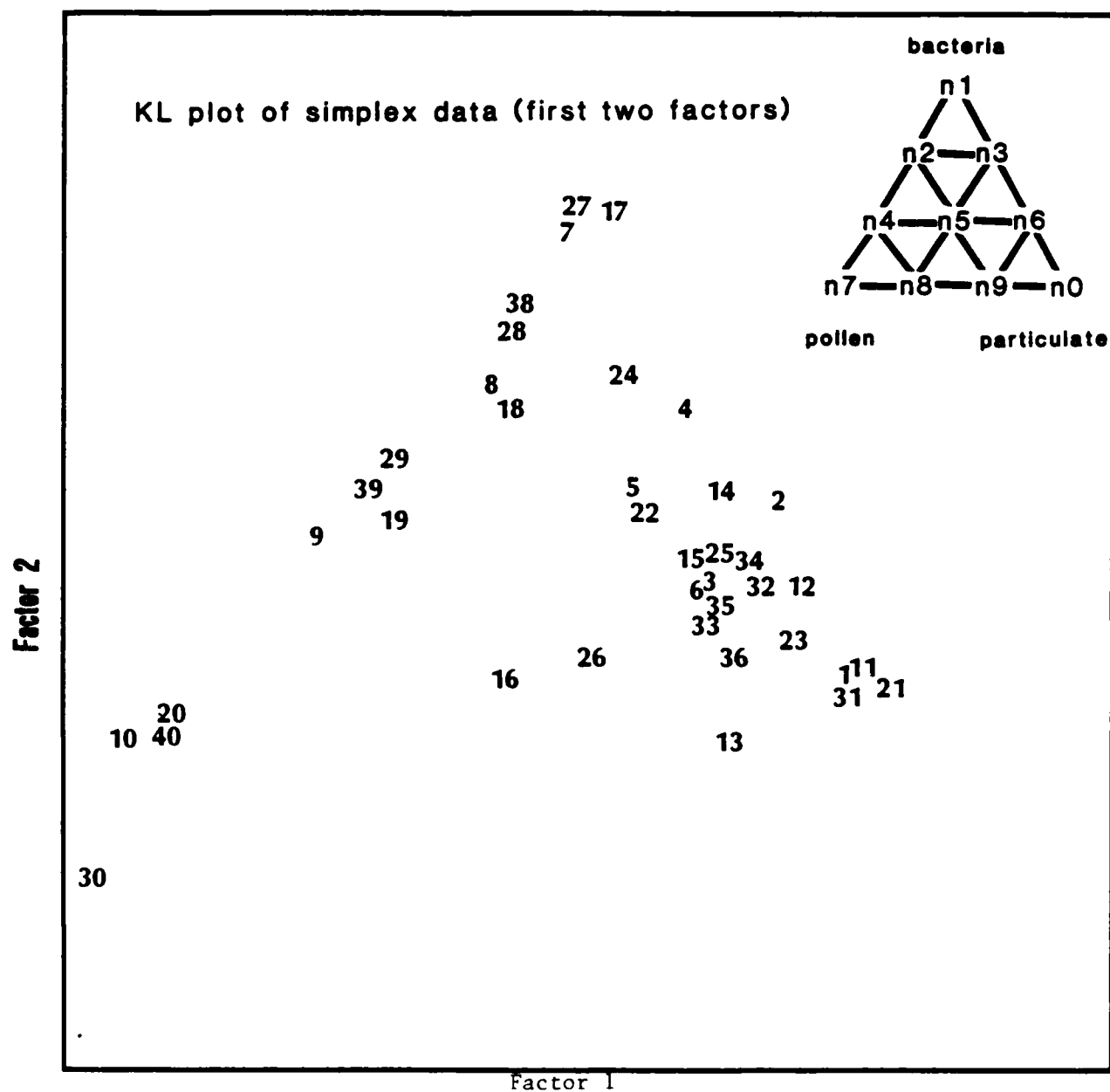


Figure 20. KL Plot of Factor 1 and Factor 2 for Simplex Data.

shows that the three pure components, samples ending in 1, 7 or 0, were quite different from each other and readily isolated. For the pollen and particulates, this implies good reproducibility, but for the bacteria it implies that the four species have traits in common which distinguish them from the other components despite the differences between species.

Figure 20, with a 120° rotation, reproduces Figure 19 quite well with respect to relative positions of the samples. This demonstrates that the amounts of the components in the mixtures can be estimated. The presence or absence of bacteria can easily be determined to the level of one part in three, and probably beyond, as indicated by the wide separation between samples containing bacteria from those without. Since this is a six component system (pollen, particulates and four bacteria), and the traits which identify the two non-bacterial components are concentrated in the first two abstract factors, the information which distinguishes the four bacteria must lie in higher factors. In Figures 21 and 22, the relevant bacteria are separated. The code letters previously established in Table 5 are used for separated species; all unseparated samples are indicated by a (.). Each of the samples containing an individual bacterium was detected in the presence of pollen and/or particulate, and was distinguished from the other three bacteria. Thus, it was possible to identify components in a set of simple mixtures and to estimate semiquantitatively the amount of each component.

Factor 6

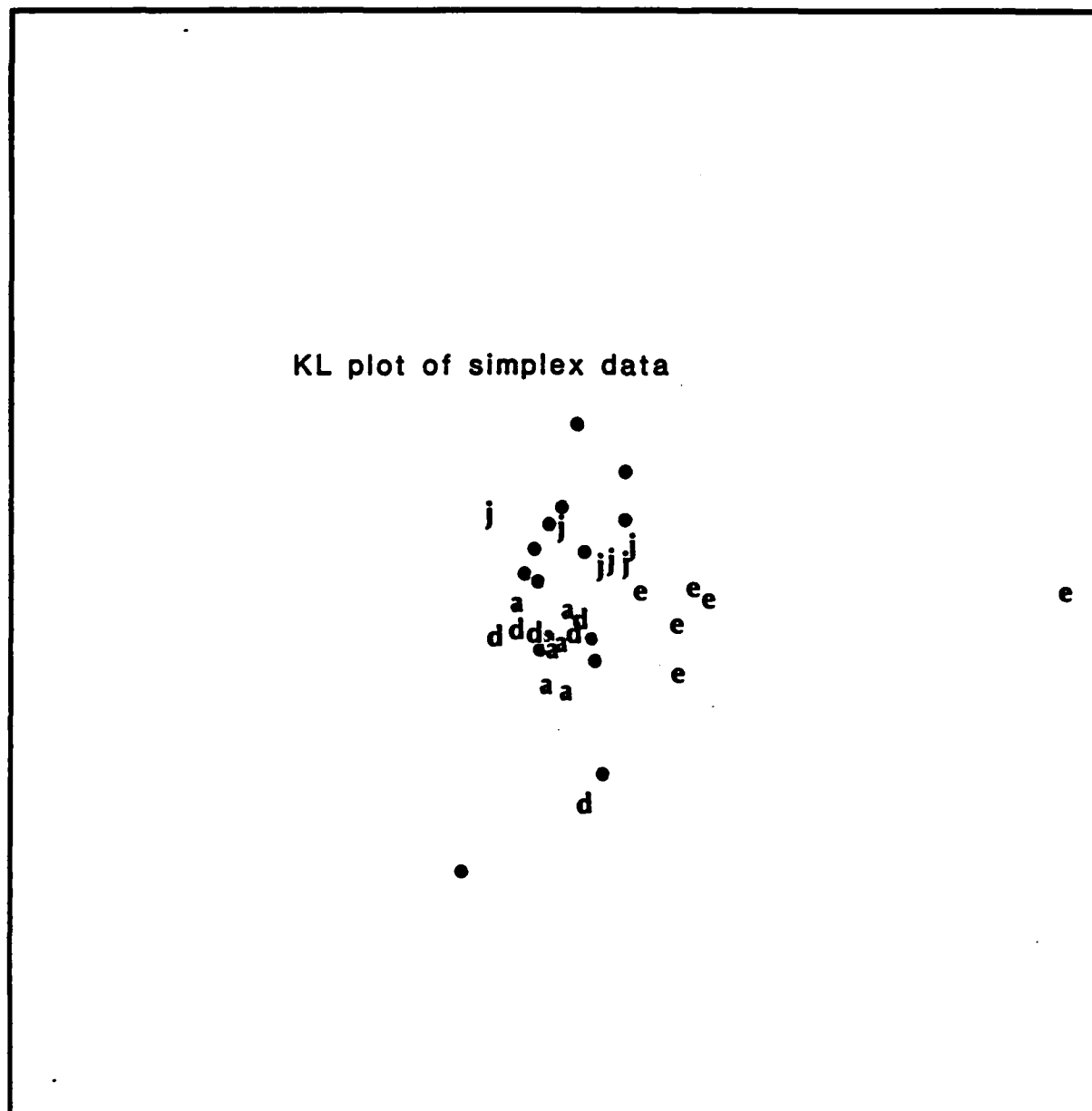


Figure 21. KL Plot of higher factors showing separations of E. coli (e).

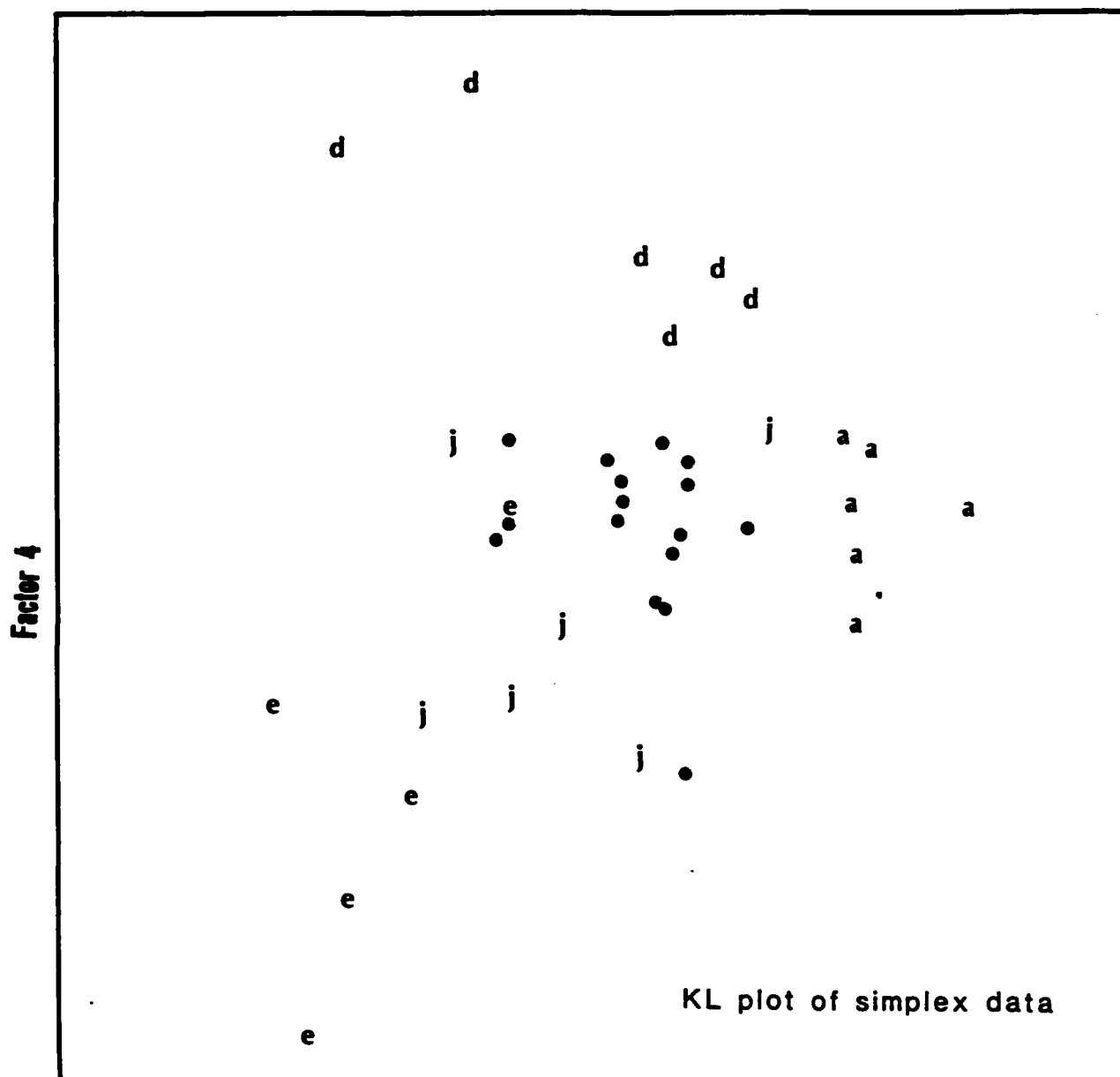


Figure 22. Addition KL Plot of bacteria in Simplex Data showing separation of P. fluorescens (d) and P. vulgaris (a).

b. Mixtures of lower concentration

The analysis of simple mixtures was quite successful, so a much more difficult data set was created involving a wider variety of bacteria and mixtures. At the same time, it was desirable to test the ability to model mixtures of somewhat lower concentrations of bacteria. In addition, because previous preparation of mixtures were made directly from live colonies, only visual (semiquantitative) estimates were made for the amount of bacteria in the simple mixtures. Therefore, in this study mixtures were prepared from suspensions of bacteria with known concentrations.

Fresh bacterial colonies were grown on trypticase soy agar from the original cultures. The bacteria on two duplicate plates were suspended in 1 mL of deionized water, transferred to culture tubes, and a second mL of deionized water was used to transfer the remaining bacteria. On the average, about 1 mL total was lost due to adhesion, incorporation into the agar and transfer losses, yielding about 3 mL of thick suspension for each bacterium. A pair of blank plates (incubated trypticase soy agar with no bacterial growth) were subjected to the same treatment. The suspensions were centrifuged followed by removal of the water and replacement with four mL of absolute ethanol. Aliquots of 0.2 mL suspension were introduced to clean tared aluminum boats and dried in a vacuum oven. From the weight of these samples, appropriate dilution factors were calculated in order to obtain 2.5 $\mu\text{g}/\mu\text{L}$ suspensions.

Suspensions of particulates of known concentration were prepared in an analogous fashion. Samples of Douglas fir smoke, red oak smoke, ABS smoke, vehicular traffic emissions, and CSM particulates were suspended in methanol by ultrasonication. They were then dried in a vacuum oven and ground into a fine powder. Approximately five microgram quantities of material were weighed into tared aluminum boats and a sufficient quantity of methanol added to yield a 5 $\mu\text{g}/\mu\text{L}$ suspension. A pollen sample was obtained as a powder and was weighed in the same manner to obtain a similar 5 $\mu\text{g}/\mu\text{L}$ suspension. Aspergillus samples were grown on agar, and were treated in the same way as the bacteria.

For every combination of one particulate and one bacterium, a single four microliter drop of bacteria was delivered to the ferromagnetic wire. Before it had dried, two four microliter drops of particulate were also added. This treatment provided approximately 50 μg total material in an intimate mixture of 1 part bacterium to four parts particulate. In addition, since fungi were expected to be a viable background source, 1:9 bacterium:mold samples of 50 μg total weight were prepared as well. Pure 50 microgram samples of each of the bacteria and particulates were also prepared.

A total of seven bacterial samples (one a blank) and seven types of particulate making a total of 49 1:4 samples were prepared. Seven 1:9 samples, seven pure bacteria and seven pure particulates were also prepared, for a total of seventy samples. These again were pyrolyzed in a previously determined random order.

The resulting data matrix was analyzed by principal components using the 16 factors with the greatest variance. These factors were subjected to graphical rotation, discriminant analysis with respect to both bacteria and particulates using SPSS (25), and the SIMCA routine in ARTHUR (19). None of these state-of-the-art modeling techniques could successfully differentiate the bacteria from the particulate background, although the particulates themselves could be readily identified and separated from each other. Figure 23 illustrates an example of the separation of the samples containing particulate material. Evidently the variation in the wide range of particulates examined in this experiment was much greater than the amount of difference between the bacteria tested. Projections, as previously described, of this data set onto models from the killing study, the growth study and the large data set could not correctly identify the bacteria.

E. CONCLUSIONS

The goals of research listed in the INTRODUCTION have been accomplished during the first year of the program. At the start of the study, the identification of bacteria by Curie-point Py-MS was limited to specimens collected and cultured under highly reproducible conditions. Also, most of the reported work up to that time had used Py/MS pattern recognition methods on a limited number of bacteria.

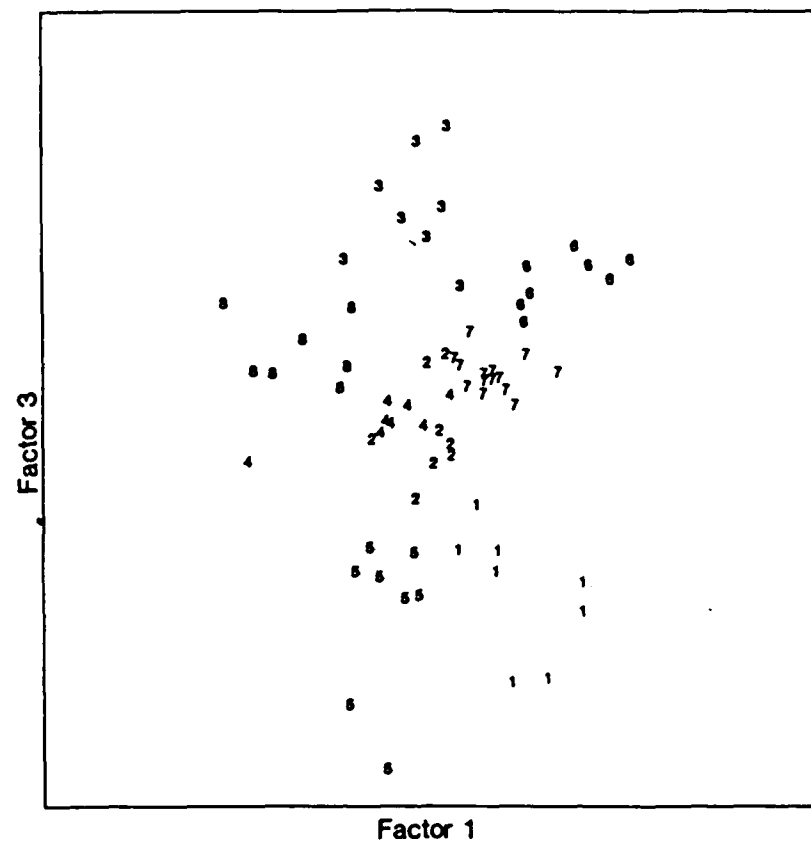


Figure 23. KL plot showing separation of major components in the complex mixtures (1 = Douglas Fir smoke, 2 = ABS smoke, 3 = Vehicular emissions, 4 = Pollen, 5 = Red Oak smoke, 6 = air particles, 7 = Aspergillus, and 8 = Bacterial Mixtures).

As a direct result of the work conducted under this contract, the understanding of the merits and limitations of single stage pyrolysis mass spectrometry for the rapid identification of targeted bacteria and molds in the presence of varying amounts and types of atmospheric background has been defined. Also, the applicability of triple quadrupole MS/MS has been assessed.

In the experiment in which a set of bacteria was grown to different stages of growth, the age of the culture was found to have little or no effect on the ability of a linear classifier to model the bacteria within the range of viability. The range of ages which might occur in a battlefield situation would probably be limited to stages of growth in which the bacteria were viable and would not affect the Py-MS technique.

Experiments to determine the effect of the method of killing showed that all bacteria studied were statistically classified without regard to a particular method. Using a K-nearest neighbors approach for classification, mixtures of bacteria in this experiment were correctly identified 75 percent of the time with only two false positives.

Triple quadrupole MS/MS experiments were performed using the parent ion and daughter ion modes. The parent ion mode provided the best data for classification, and characteristic parent-daughter ion pairs were found for all three of the bacteria tested. A methodology was developed for isolating the parent-daughter pairs for other bacteria. Since the parent-daughter pairs were unique for the bacteria tested, it appears that the

technique could provide a rapid, accurate and potentially sensitive method for detecting bacteria from mixtures.

Mixture experiments were performed for a set of four species of bacteria in the presence of varying contributions from two potential contaminants. The presence of bacteria was readily detected, and, once detected, the species of bacteria present could be determined. Mixture experiments involving more complex backgrounds (seven completely different contaminants with lower concentrations of seven different species of bacteria) were not successful, (approximately 30 percent reliably) indicating that simple, one stage pyrolysis-MS will require a previous separation, characterization of the background, or exploitation of changes in the background load relative to the bacteria.

Projection of test sets of bacteria, treated by the computer as unknowns, were classified reasonably well in the one-stage pyrolysis-MS experiments; however, the difficulty in characterizing bacteria in low concentrations when a highly variable assortment of contaminants was present strongly suggests that the triple-quadrupole MS/MS approach is necessary for identifying bacteria in such mixtures.

The failure of the Py/MS pattern recognition approach to successfully identify bacteria in mixtures of lower concentration and greater diversity was disappointing. The same approach gave positive results with a less complex comparison set in the analysis of two or three component mixtures. This implies that the limits of detection for bacteria in low concentrations relative to background lies somewhere between these two extremes.

Several possibilities can be suggested for modifying the single stage Py-MS pattern recognition approach. As indicated by the simple mixtures, a well-characterized background is not a problem - difficulties become apparent as the background becomes highly variable. Therefore, one approach might be to sample the background without bacteria, and then construct a training set based on the known background. Another possibility would be a prior separation step such as a cyclone separator to remove pollen and larger particles. This would increase the amount of bacteria relative to background based on size. Direct chemical ionization has been found to provide more specific pyrolysis mass spectra than electron impact ionization (26).

The results from this study indicate that some form of tandem mass spectrometry will be necessary in future development of Py-MS for bacterial identification. Tandem mass spectrometry will add selectivity to the experiment while at the same time not affect the positive advantages of Py-MS. Research in the area of Py-MS/MS for bacterial identification should remain a high priority for the Army.

REFERENCES

1. Whiton, R.S., M.D. Walla and Stephen L. Morgan, Private Communication (1986).
2. Eudy, L.W., M.D. Walla, J.R. Hudson and S.L. Morgan, J. Anal. Appl. Pyrol., 7, 231-247 (1985).
3. Eudy, L.W., M.D. Walla, S.L. Morgan and A. Fox, Analyst, 10, 381-385 (1985).
4. Walla, M.D., P.Y. Lau, S.L. Morgan, A. Fox and A. Brown, J. Chromatogr., 288, 399-413 (1984).
5. Reiner, E. and W.H. Ewing, Nature, 217, 191-194 (1968).
6. Gutteridge, C.S. and J.R. Norris, Appl. Environ. Microbiology, 40, 462-465 (1980).
7. Soderstrom, B., S. Wold and G. Blomquist, J. Gen. Microbiol., 128, 1773-1784 (1982).
8. Bayer, F.L. and S.L. Morgan, In "Pyrolysis and GC in Polymer Analysis" S.A. Liebman and E.J. Levy, Eds., Marcel Dekker, NY, 1985, p. 277-337.
9. Meuzelaar, H.L.C., J. Haverkamp and F.D. Hileman, "Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials", Elsevier, Amsterdam, 1982.
10. Irwin, W.J., J. Anal. Appl. Pyrol., 1, 89-122 (1979).
11. Risby, T.H. and A.L. Yergey, J. Phys. Chem., 80, 2839-2845 (1976).

12. Shute, L.A., C.S. Gutteridge, J.R. Norris and R.C.W. Berkely, J. Gen. Microbiol., 130, 343-355 (1984).
13. Kistemaker, P.G., H.L.C. Meuzelaar and M.A. Posthumus, In "New Approaches to the Identification of Microorganisms", C-G. Heden and T. Illeni (Eds.), John Wiley, NY, 1975, p. 179-191.
14. Kajioka, R. and P.W. Tang, J. Anal. App. Pyrol., 6, 59-68 (1984).
15. Sinha, M.P., R.M. Platz, U.L. Vilker and S.R. Friedlander, Int. J. Mass Spectrum & Ion Proc., 57, 125-136 (1984).
16. de Leeuw, J.W., W.L. Maters, D.v.d. Meent and J.J. Boon, Anal. Chem., 49, 1881-1884 (1977).
17. Durfee, S.L. and K.J. Voorhees, Anal. Chem., 57, 2378-2383 (1985).
18. Rozett, R.W. and E.M. Peterson, Anal. Chem., 47, 1301-1308 (1975).
19. Harper, A.M., D.L. Duewer, B.R. Kowalski, J.L. Fasching, In "Chemometrics: Theory and Applications", Kowalski, B.R., Ed., 1977, ACS Symp. Ser. No. 52, p. 14.
20. Windig, W., J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55, 81-86 (1983).

21. Montgomery, D.C. "Design and Analysis of Experiments," John Wiley and Sons, New York, 1976.
22. Windig, W. and H.L.C. Meuzelaar, Anal. Chem., 56, 2297-2303 (1984).
23. Yost, R.A. and G.G. Enke, Anal. Chem., 51, 1251A-1260A (1979).
24. McLafferty, F., "Tandem Mass Spectrometry," Wiley, New York, 1983.
25. Users Guide, 2nd Edition, IBM Corp., 444 North Michigan Ave., Chicago, IL 60611.
26. Harrison, A.G., "Chemical Analysis by Mass Spectrometry," CRC Press, Boca Raton, FL, 1981.

END

3-87

Dtic